

Our Ref.: CL001006-CIP

U.S. PATENT APPLICATION

Inventor(s): Rhonda C. BRANDON
Valentina DI FRANCESCO
Ellen M. BEASLEY

Invention: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID
MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND
USES THEREOF

CELERA GENOMICS CORPORATION.
45 WEST GUDE DR., C2-4#20
ROCKVILLE, MD 20850
(240) 453-3067
Fax (240)-453-3084

SPECIFICATION

**ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID
MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
AND USES THEREOF**

5 **RELATED APPLICATIONS**

The present application is a Continuation-In-Part of US Serial No. 09/730,002,
filed December 8, 2000 (Atty. Docket CL001066).

FIELD OF THE INVENTION

10 The present invention is in the field of transporter proteins that are related to the
gamma-aminobutyric-acid receptor subfamily, recombinant DNA molecules, and protein
production. The present invention specifically provides novel peptides and proteins that
effect ligand transport and nucleic acid molecules encoding such peptide and protein
molecules, all of which are useful in the development of human therapeutics and
diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Transporters

15 Transporter proteins regulate many different functions of a cell, including cell
proliferation, differentiation, and signaling processes, by regulating the flow of molecules
such as ions and macromolecules, into and out of cells. Transporters are found in the
plasma membranes of virtually every cell in eukaryotic organisms. Transporters mediate
a variety of cellular functions including regulation of membrane potentials and absorption
and secretion of molecules and ion across cell membranes. When present in intracellular
membranes of the Golgi apparatus and endocytic vesicles, transporters, such as chloride
channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev.

20 Physiol. 50:111-122.

25 Transporters are generally classified by structure and the type of mode of action.
In addition, transporters are sometimes classified by the molecule type that is transported,
for example, sugar transporters, chlorine channels, potassium channels, etc. There may

be many classes of channels for transporting a single type of molecule (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters: Receptor and transporter nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 (1997) and <http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html>.

5 The following general classification scheme is known in the art and is followed in the present discoveries.

Channel-type transporters. Transmembrane channel proteins of this class are ubiquitously found in the membranes of all types of organisms from bacteria to higher eukaryotes. Transport systems of this type catalyze facilitated diffusion (by an energy-independent process) by passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism. These channel proteins usually consist largely of α -helical spanners, although β -strands may also be present and may even comprise the channel. However, outer membrane porin-type channel proteins are excluded from this class and are instead included in class 9.

Carrier-type transporters. Transport systems are included in this class if they utilize a carrier-mediated process to catalyze uniport (a single species is transported by facilitated diffusion), antiport (two or more species are transported in opposite directions in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy) and/or symport (two or more species are transported together in the same direction in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy).

Pyrophosphate bond hydrolysis-driven active transporters. Transport systems are included in this class if they hydrolyze pyrophosphate or the terminal pyrophosphate bond in ATP or another nucleoside triphosphate to drive the active uptake and/or extrusion of a solute or solutes. The transport protein may or may not be transiently phosphorylated, but the substrate is not phosphorylated.

PEP-dependent, phosphoryl transfer-driven group translocators. Transport systems of the bacterial phosphoenolpyruvate:sugar phosphotransferase system are included in this class. The product of the reaction, derived from extracellular sugar, is a cytoplasmic sugar-phosphate.

Decarboxylation-driven active transporters. Transport systems that drive solute (e.g., ion) uptake or extrusion by decarboxylation of a cytoplasmic substrate are included in this class.

Oxidoreduction-driven active transporters. Transport systems that drive transport of a solute (e.g., an ion) energized by the flow of electrons from a reduced substrate to an oxidized substrate are included in this class.

Light-driven active transporters. Transport systems that utilize light energy to drive transport of a solute (e.g., an ion) are included in this class.

Mechanically-driven active transporters. Transport systems are included in this class if they drive movement of a cell or organelle by allowing the flow of ions (or other solutes) through the membrane down their electrochemical gradients.

Outer-membrane porins (of β -structure). These proteins form transmembrane pores or channels that usually allow the energy independent passage of solutes across a membrane. The transmembrane portions of these proteins consist exclusively of β -strands that form a β -barrel. These porin-type proteins are found in the outer membranes of Gram-negative bacteria, mitochondria and eukaryotic plastids.

Methyltransferase-driven active transporters. A single characterized protein currently falls into this category, the Na^+ -transporting methyltetrahydromethanopterin:coenzyme M methyltransferase.

Non-ribosome-synthesized channel-forming peptides or peptide-like molecules. These molecules, usually chains of L- and D-amino acids as well as other small molecular building blocks such as lactate, form oligomeric transmembrane ion channels. Voltage may induce channel formation by promoting assembly of the transmembrane channel. These peptides are often made by bacteria and fungi as agents of biological warfare.

Non-Proteinaceous Transport Complexes. Ion conducting substances in biological membranes that do not consist of or are not derived from proteins or peptides fall into this category.

Functionally characterized transporters for which sequence data are lacking. Transporters of particular physiological significance will be included in this category even though a family assignment cannot be made.

Putative transporters in which no family member is an established transporter. Putative transport protein families are grouped under this number and will either be classified elsewhere when the transport function of a member becomes established, or will be eliminated from the TC classification system if the proposed transport function is disproven. These families include a member or members for which a transport function has been suggested, but evidence for such a function is not yet compelling.

Auxiliary transport proteins. Proteins that in some way facilitate transport across one or more biological membranes but do not themselves participate directly in transport are included in this class. These proteins always function in conjunction with one or more transport proteins. They may provide a function connected with energy coupling to transport, play a structural role in complex formation or serve a regulatory function.

Transporters of unknown classification. Transport protein families of unknown classification are grouped under this number and will be classified elsewhere when the transport process and energy coupling mechanism are characterized. These families include at least one member for which a transport function has been established, but either the mode of transport or the energy coupling mechanism is not known.

Ion channels

An important type of transporter is the ion channel. Ion channels regulate many different cell proliferation, differentiation, and signaling processes by regulating the flow of ions into and out of cells. Ion channels are found in the plasma membranes of virtually every cell in eukaryotic organisms. Ion channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, ion channels, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) *Annu. Rev. Physiol.* 50:111-122.

Ion channels are generally classified by structure and the type of mode of action. For example, extracellular ligand gated channels (ELGs) are comprised of five polypeptide subunits, with each subunit having 4 membrane spanning domains, and are activated by the binding of an extracellular ligand to the channel. In addition, channels are sometimes classified by the ion type that is transported, for example, chlorine channels, potassium channels, etc. There may be many classes of channels for

transporting a single type of ion (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters (1997). Receptor and ion channel nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 and <http://www-biology.ucsd.edu/~msaier/transport/toc.html>).

There are many types of ion channels based on structure. For example, many ion channels fall within one of the following groups: extracellular ligand-gated channels (ELG), intracellular ligand-gated channels (ILG), inward rectifying channels (INR), intercellular (gap junction) channels, and voltage gated channels (VIC). There are additionally recognized other channel families based on ion-type transported, cellular location and drug sensitivity. Detailed information on each of these, their activity, ligand type, ion type, disease association, drugability, and other information pertinent to the present invention, is well known in the art.

Extracellular ligand-gated channels, ELGs, are generally comprised of five polypeptide subunits, Unwin, N. (1993), Cell 72: 31-41; Unwin, N. (1995), Nature 373: 37-43; Hucho, F., et al., (1996) J. Neurochem. 66: 1781-1792; Hucho, F., et al., (1996) Eur. J. Biochem. 239: 539-557; Alexander, S.P.H. and J.A. Peters (1997), Trends Pharmacol. Sci., Elsevier, pp. 4-6; 36-40; 42-44; and Xue, H. (1998) J. Mol. Evol. 47: 323-333. Each subunit has 4 membrane spanning regions: this serves as a means of identifying other members of the ELG family of proteins. ELG bind a ligand and in response modulate the flow of ions. Examples of ELG include most members of the neurotransmitter-receptor family of proteins, e.g., GABAI receptors. Other members of this family of ion channels include glycine receptors, ryandyne receptors, and ligand gated calcium channels.

The Voltage-gated Ion Channel (VIC) Superfamily

Proteins of the VIC family are ion-selective channel proteins found in a wide range of bacteria, archaea and eukaryotes Hille, B. (1992), Chapter 9: Structure of channel proteins; Chapter 20: Evolution and diversity. In: Ionic Channels of Excitable Membranes, 2nd Ed., Sinaur Assoc. Inc., Pubs., Sunderland, Massachusetts; Sigworth, F.J. (1993), Quart. Rev. Biophys. 27: 1-40; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Alexander, S.P.H. et al., (1997), Trends Pharmacol. Sci., Elsevier, pp. 76-84; Jan, L.Y. et al., (1997), Annu. Rev. Neurosci. 20: 91-123; Doyle, D.A., et al., (1998)

Science 280: 69-77; Terlau, H. and W. Stühmer (1998), *Naturwissenschaften* 85: 437-444. They are often homo- or heterooligomeric structures with several dissimilar subunits (e.g., $\alpha_1\alpha_2\gamma\delta$ Ca^{2+} channels, $\alpha_1\beta_2$ Na^+ channels or $(\alpha)_4\beta$ K^+ channels), but the channel and the primary receptor is usually associated with the α (or α_1) subunit.

- 5 Functionally characterized members are specific for K^+ , Na^+ or Ca^{2+} . The K^+ channels usually consist of homotetrameric structures with each α -subunit possessing six transmembrane spanners (TMSs). The α_1 and α subunits of the Ca^{2+} and Na^+ channels, respectively, are about four times as large and possess 4 units, each with 6 TMSs separated by a hydrophilic loop, for a total of 24 TMSs. These large channel proteins form heterotetra-unit structures equivalent to the homotetrameric structures of most K^+ channels. All four units of the Ca^{2+} and Na^+ channels are homologous to the single unit in the homotetrameric K^+ channels. Ion flux via the eukaryotic channels is generally controlled by the transmembrane electrical potential (hence the designation, voltage-sensitive) although some are controlled by ligand or receptor binding.

Several putative K^+ -selective channel proteins of the VIC family have been identified in prokaryotes. The structure of one of them, the KcsA K^+ channel of *Streptomyces lividans*, has been solved to 3.2 Å resolution. The protein possesses four identical subunits, each with two transmembrane helices, arranged in the shape of an inverted teepee or cone. The cone cradles the "selectivity filter" P domain in its outer end.

- 20 The narrow selectivity filter is only 12 Å long, whereas the remainder of the channel is wider and lined with hydrophobic residues. A large water-filled cavity and helix dipoles stabilize K^+ in the pore. The selectivity filter has two bound K^+ ions about 7.5 Å apart from each other. Ion conduction is proposed to result from a balance of electrostatic attractive and repulsive forces.

- 25 In eukaryotes, each VIC family channel type has several subtypes based on pharmacological and electrophysiological data. Thus, there are five types of Ca^{2+} channels (L, N, P, Q and T). There are at least ten types of K^+ channels, each responding in different ways to different stimuli: voltage-sensitive [K_A, K_V, K_{VR}, K_{VS} and K_{SR}], Ca^{2+} -sensitive [BK_{Ca}, IK_{Ca} and SK_{Ca}] and receptor-coupled [K_M and K_{ACH}]. There are at least
30 six types of Na^+ channels (I, II, III, μ 1, H1 and PN3). Tetrameric channels from both prokaryotic and eukaryotic organisms are known in which each α -subunit possesses 2

TMSs rather than 6, and these two TMSs are homologous to TMSs 5 and 6 of the six TMS unit found in the voltage-sensitive channel proteins. KcsA of *S. lividans* is an example of such a 2 TMS channel protein. These channels may include the K_{Na} (Na^+ -activated) and K_{Vol} (cell volume-sensitive) K^+ channels, as well as distantly related channels such as the Tok1 K^+ channel of yeast, the TWIK-1 inward rectifier K^+ channel of the mouse and the TREK-1 K^+ channel of the mouse. Because of insufficient sequence similarity with proteins of the VIC family, inward rectifier K^+ IRK channels (ATP-regulated; G-protein-activated) which possess a P domain and two flanking TMSs are placed in a distinct family. However, substantial sequence similarity in the P region suggests that they are homologous. The b, g and d subunits of VIC family members, when present, frequently play regulatory roles in channel activation/deactivation.

The Epithelial Na^+ Channel (ENaC) Family

The ENaC family consists of over twenty-four sequenced proteins (Canessa, C.M., et al., (1994), Nature 367: 463-467, Le, T. and M.H. Saier, Jr. (1996), Mol. Membr. Biol. 13: 149-157; Garty, H. and L.G. Palmer (1997), Physiol. Rev. 77: 359-396; Waldmann, R., et al., (1997), Nature 386: 173-177; Darboux, I., et al., (1998), J. Biol. Chem. 273: 9424-9429; Firsov, D., et al., (1998), EMBO J. 17: 344-352; Horisberger, J.-D. (1998). Curr. Opin. Struc. Biol. 10: 443-449). All are from animals with no recognizable homologues in other eukaryotes or bacteria. The vertebrate ENaC proteins from epithelial cells cluster tightly together on the phylogenetic tree: voltage-insensitive ENaC homologues are also found in the brain. Eleven sequenced *C. elegans* proteins, including the degenerins, are distantly related to the vertebrate proteins as well as to each other. At least some of these proteins form part of a mechano-transducing complex for touch sensitivity. The homologous *Helix aspersa* (FMRF-amide)-activated Na^+ channel is the first peptide neurotransmitter-gated ionotropic receptor to be sequenced.

Protein members of this family all exhibit the same apparent topology, each with N- and C-termini on the inside of the cell, two amphipathic transmembrane spanning segments, and a large extracellular loop. The extracellular domains contain numerous highly conserved cysteine residues. They are proposed to serve a receptor function.

Mammalian ENaC is important for the maintenance of Na^+ balance and the regulation of blood pressure. Three homologous ENaC subunits, alpha, beta, and gamma,

have been shown to assemble to form the highly Na^+ -selective channel. The stoichiometry of the three subunits is $\alpha_2\beta_1\gamma_1$ in a heterotetrameric architecture.

The Glutamate-gated Ion Channel (GIC) Family of Neurotransmitter Receptors

Members of the GIC family are heteropentameric complexes in which each of the 5 subunits is of 800-1000 amino acid residues in length (Nakanishi, N., et al, (1990), Neuron 5: 569-581; Unwin, N. (1993), Cell 72: 31-41; Alexander, S.P.H. and J.A. Peters (1997) Trends Pharmacol. Sci., Elsevier, pp. 36-40). These subunits may span the membrane three or five times as putative α -helices with the N-termini (the glutamate-binding domains) localized extracellularly and the C-termini localized cytoplasmically. They may be distantly related to the ligand-gated ion channels, and if so, they may possess substantial β -structure in their transmembrane regions. However, homology between these two families cannot be established on the basis of sequence comparisons alone. The subunits fall into six subfamilies: a, b, g, d, e and z.

The GIC channels are divided into three types: (1) α -amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA)-, (2) kainate- and (3) N-methyl-D-aspartate (NMDA)-selective glutamate receptors. Subunits of the AMPA and kainate classes exhibit 35-40% identity with each other while subunits of the NMDA receptors exhibit 22-24% identity with the former subunits. They possess large N-terminal, extracellular glutamate-binding domains that are homologous to the periplasmic glutamine and glutamate receptors of ABC-type uptake permeases of Gram-negative bacteria. All known members of the GIC family are from animals. The different channel (receptor) types exhibit distinct ion selectivities and conductance properties. The NMDA-selective large conductance channels are highly permeable to monovalent cations and Ca^{2+} . The AMPA- and kainate-selective ion channels are permeable primarily to monovalent cations with only low permeability to Ca^{2+} .

The Chloride Channel (ClC) Family

The ClC family is a large family consisting of dozens of sequenced proteins derived from Gram-negative and Gram-positive bacteria, cyanobacteria, archaea, yeast, plants and animals (Steinmeyer, K., et al., (1991), Nature 354: 301-304; Uchida, S., et al.,

(1993), J. Biol. Chem. 268: 3821-3824; Huang, M.-E., et al., (1994), J. Mol. Biol. 242: 595-598; Kawasaki, M., et al., (1994), Neuron 12: 597-604; Fisher, W.E., et al., (1995), Genomics. 29:598-606; and Foskett, J.K. (1998), Annu. Rev. Physiol. 60: 689-717).

These proteins are essentially ubiquitous, although they are not encoded within genomes of *Haemophilus influenzae*, *Mycoplasma genitalium*, and *Mycoplasma pneumoniae*.

Sequenced proteins vary in size from 395 amino acid residues (*M. jannaschii*) to 988 residues (man). Several organisms contain multiple ClC family paralogues. For example, *Synechocystis* has two paralogues, one of 451 residues in length and the other of 899 residues. *Arabidopsis thaliana* has at least four sequenced paralogues, (775-792

residues), humans also have at least five paralogues (820-988 residues), and *C. elegans* also has at least five (810-950 residues). There are nine known members in mammals, and mutations in three of the corresponding genes cause human diseases. *E. coli*, *Methanococcus jannaschii* and *Saccharomyces cerevisiae* only have one ClC family member each. With the exception of the larger *Synechocystis* paralogue, all bacterial proteins are small (395-492 residues) while all eukaryotic proteins are larger (687-988 residues). These proteins exhibit 10-12 putative transmembrane α -helical spanners (TMSs) and appear to be present in the membrane as homodimers. While one member of the family, *Torpedo* ClC-O, has been reported to have two channels, one per subunit, others are believed to have just one.

All functionally characterized members of the ClC family transport chloride, some in a voltage-regulated process. These channels serve a variety of physiological functions (cell volume regulation; membrane potential stabilization; signal transduction; transepithelial transport, etc.). Different homologues in humans exhibit differing anion selectivities, i.e., ClC4 and ClC5 share a $\text{NO}_3^- > \text{Cl}^- > \text{Br}^- > \text{I}^-$ conductance sequence, while ClC3 has an $\text{I}^- > \text{Cl}^-$ selectivity. The ClC4 and ClC5 channels and others exhibit outward rectifying currents with currents only at voltages more positive than +20mV.

Animal Inward Rectifier K^+ Channel (IRK-C) Family

IRK channels possess the "minimal channel-forming structure" with only a P domain, characteristic of the channel proteins of the VIC family, and two flanking transmembrane spanners (Shuck, M.E., et al., (1994), J. Biol. Chem. 269: 24261-24270; Ashen, M.D., et al., (1995), Am. J. Physiol. 268: H506-H511; Salkoff, L. and T. Jegla

(1995), Neuron 15: 489-492; Aguilar-Bryan, L., et al., (1998), Physiol. Rev. 78: 227-245; Ruknudin, A., et al., (1998), J. Biol. Chem. 273: 14165-14171). They may exist in the membrane as homo- or heterooligomers. They have a greater tendency to let K^+ flow into the cell than out. Voltage-dependence may be regulated by external K^+ , by internal Mg^{2+} , by internal ATP and/or by G-proteins. The P domains of IRK channels exhibit limited sequence similarity to those of the VIC family, but this sequence similarity is insufficient to establish homology. Inward rectifiers play a role in setting cellular membrane potentials, and the closing of these channels upon depolarization permits the occurrence of long duration action potentials with a plateau phase. Inward rectifiers lack the intrinsic voltage sensing helices found in VIC family channels. In a few cases, those of Kir1.1a and Kir6.2, for example, direct interaction with a member of the ABC superfamily has been proposed to confer unique functional and regulatory properties to the heteromeric complex, including sensitivity to ATP. The SUR1 sulfonylurea receptor (spQ09428) is the ABC protein that regulates the Kir6.2 channel in response to ATP, and CFTR may regulate Kir1.1a. Mutations in SUR1 are the cause of familial persistent hyperinsulinemic hypoglycemia in infancy (PHHI), an autosomal recessive disorder characterized by unregulated insulin secretion in the pancreas.

ATP-gated Cation Channel (ACC) Family

Members of the ACC family (also called P2X receptors) respond to ATP, a functional neurotransmitter released by exocytosis from many types of neurons (North, R.A. (1996), Curr. Opin. Cell Biol. 8: 474-483; Soto, F., M. Garcia-Guzman and W. Stühmer (1997), J. Membr. Biol. 160: 91-100). They have been placed into seven groups (P2X₁ - P2X₇) based on their pharmacological properties. These channels, which function at neuron-neuron and neuron-smooth muscle junctions, may play roles in the control of blood pressure and pain sensation. They may also function in lymphocyte and platelet physiology. They are found only in animals.

The proteins of the ACC family are quite similar in sequence (>35% identity), but they possess 380-1000 amino acid residues per subunit with variability in length localized primarily to the C-terminal domains. They possess two transmembrane spanners, one about 30-50 residues from their N-termini, the other near residues 320-340. The extracellular receptor domains between these two spanners (of about 270 residues) are

well conserved with numerous conserved glycyl and cysteyl residues. The hydrophilic C-termini vary in length from 25 to 240 residues. They resemble the topologically similar epithelial Na^+ channel (ENaC) proteins in possessing (a) N- and C-termini localized intracellularly, (b) two putative transmembrane spanners, (c) a large extracellular loop domain, and (d) many conserved extracellular cysteyl residues. ACC family members are, however, not demonstrably homologous with them. ACC channels are probably hetero- or homomultimers and transport small monovalent cations (Me^+). Some also transport Ca^{2+} ; a few also transport small metabolites.

The Ryanodine-Inositol 1,4,5-triphosphate Receptor Ca^{2+} Channel (RIR-CaC)

Family

Ryanodine (Ry)-sensitive and inositol 1,4,5-triphosphate (IP_3)-sensitive Ca^{2+} -release channels function in the release of Ca^{2+} from intracellular storage sites in animal cells and thereby regulate various Ca^{2+} -dependent physiological processes (Hasan, G. et al., (1992) Development 116: 967-975; Michikawa, T., et al., (1994), J. Biol. Chem. 269: 9184-9189; Tunwell, R.E.A., (1996), Biochem. J. 318: 477-487; Lee, A.G. (1996) *Biomembranes*, Vol. 6, Transmembrane Receptors and Channels (A.G. Lee, ed.), JAI Press, Denver, CO., pp 291-326; Mikoshiba, K., et al., (1996) J. Biochem. Biomem. 6: 273-289). Ry receptors occur primarily in muscle cell sarcoplasmic reticular (SR) membranes, and IP_3 receptors occur primarily in brain cell endoplasmic reticular (ER) membranes where they effect release of Ca^{2+} into the cytoplasm upon activation (opening) of the channel.

The Ry receptors are activated as a result of the activity of dihydropyridine-sensitive Ca^{2+} channels. The latter are members of the voltage-sensitive ion channel (VIC) family. Dihydropyridine-sensitive channels are present in the T-tubular systems of muscle tissues.

Ry receptors are homotetrameric complexes with each subunit exhibiting a molecular size of over 500,000 daltons (about 5,000 amino acid residues). They possess C-terminal domains with six putative transmembrane α -helical spanners (TMSs). Putative pore-forming sequences occur between the fifth and sixth TMSs as suggested for members of the VIC family. The large N-terminal hydrophilic domains and the small C-terminal hydrophilic domains are localized to the cytoplasm. Low resolution 3-

dimensional structural data are available. Mammals possess at least three isoforms that probably arose by gene duplication and divergence before divergence of the mammalian species. Homologues are present in humans and *Caenorabditis elegans*.

IP₃ receptors resemble Ry receptors in many respects. (1) They are homotetrameric complexes with each subunit exhibiting a molecular size of over 300,000 daltons (about 2,700 amino acid residues). (2) They possess C-terminal channel domains that are homologous to those of the Ry receptors. (3) The channel domains possess six putative TMSs and a putative channel lining region between TMSs 5 and 6. (4) Both the large N-terminal domains and the smaller C-terminal tails face the cytoplasm. (5) They possess covalently linked carbohydrate on extracytoplasmic loops of the channel domains. (6) They have three currently recognized isoforms (types 1, 2, and 3) in mammals which are subject to differential regulation and have different tissue distributions.

IP₃ receptors possess three domains: N-terminal IP₃-binding domains, central coupling or regulatory domains and C-terminal channel domains. Channels are activated by IP₃ binding, and like the Ry receptors, the activities of the IP₃ receptor channels are regulated by phosphorylation of the regulatory domains, catalyzed by various protein kinases. They predominate in the endoplasmic reticular membranes of various cell types in the brain but have also been found in the plasma membranes of some nerve cells derived from a variety of tissues.

The channel domains of the Ry and IP₃ receptors comprise a coherent family that in spite of apparent structural similarities, do not show appreciable sequence similarity of the proteins of the VIC family. The Ry receptors and the IP₃ receptors cluster separately on the RIR-CaC family tree. They both have homologues in *Drosophila*. Based on the phylogenetic tree for the family, the family probably evolved in the following sequence: (1) A gene duplication event occurred that gave rise to Ry and IP₃ receptors in invertebrates. (2) Vertebrates evolved from invertebrates. (3) The three isoforms of each receptor arose as a result of two distinct gene duplication events. (4) These isoforms were transmitted to mammals before divergence of the mammalian species.

The Organellar Chloride Channel (O-ClC) Family

Proteins of the O-ClC family are voltage-sensitive chloride channels found in intracellular membranes but not the plasma membranes of animal cells (Landry, D, et al., (1993), J. Biol. Chem. 268: 14948-14955; Valenzuela, Set al., (1997), J. Biol. Chem. 272: 12575-12582; and Duncan, R.R., et al., (1997), J. Biol. Chem. 272: 23880-23886).

They are found in human nuclear membranes, and the bovine protein targets to the microsomes, but not the plasma membrane, when expressed in *Xenopus laevis* oocytes. These proteins are thought to function in the regulation of the membrane potential and in transepithelial ion absorption and secretion in the kidney. They possess two putative transmembrane α -helical spanners (TMSs) with cytoplasmic N- and C-termini and a large luminal loop that may be glycosylated. The bovine protein is 437 amino acid residues in length and has the two putative TMSs at positions 223-239 and 367-385. The human nuclear protein is much smaller (241 residues). A *C. elegans* homologue is 260 residues long.

Gamma-Aminobutyric acidA receptors

Gamma-Aminobutyric acid A (GABAA) receptors are multisubunit ligand-gated ion channels which mediate neuronal inhibition by GABA and are composed of at least four subunit types (alpha, beta, gamma, and delta). The gamma 2-subunit appears to be essential for benzodiazepine modulation of GABAA receptor function. The gamma-aminobutyric acid receptors are the major inhibitory neurotransmitter receptors in the brain and the site of action of a number of important pharmacologic agents including barbiturates, benzodiazepines, and ethanol. The gamma-1 and gamma-2 subunits are important in mediating responses to benzodiazepines, and a splicing variant of the gamma-2 subunit, gamma-2L, is necessary for ethanol actions on the receptor, raising the possibility that the gamma-2 gene may be involved in human genetic predisposition to the development of alcoholism. Wilcox et al. (1992) assigned genes encoding the gamma-1 and gamma-2 subunits of the GABA(A) receptor to chromosomes 4 and 5, respectively, by PCR amplification of human-specific products from human-hamster somatic cell hybrid DNAs. Using panels of chromosome-specific natural deletion hybrids, Wilcox et al. (1992) further localized the gamma-1 gene (GABRG1) to 4p14-

q21.1 and the gamma-2 gene (GABRG2) to 5q31.1-q33.2. These data suggested that the GABRG1 gene may be clustered with the previously mapped GABRA2 and GABRB1 gene on chromosome 4 and that the GABRG2 gene may be close to the previously localized GABRA1 gene on chromosome 5. To test this further, the GABRA1 gene was mapped using the chromosome 5 deletion hybrids and shown to be within the same region as the GABRG2 gene, 5q31.1-q33.2. By means of a PCR-based screening strategy, a 450-kb human genomic YAC clone containing both the GABRA1 and the GABRG2 genes was isolated. Pulsed field gel restriction mapping of this YAC indicated that the 2 genes are within 200 kb of each other.

GABA-A ligand-gated channels complex selectively with dopamine D5 receptors through the direct binding of the D5 carboxy-terminal domain with the second intracellular loop of the GABA-A gamma-2 (short) receptor subunit. This physical association enables mutual inhibitory functional interactions between these receptor systems. Therefore it involves in an unknown signal transduction mechanism whereby subtype-selective G protein-coupled receptors dynamically regulate synaptic strength independently of classically defined second-messenger systems, and suggest a possible framework in which to view these receptor systems in the maintenance of psychomotor disease states, particularly schizophrenia.

For a review related to the GABA-A receptor, see reference of Ymer *et al.*, *EMBO J* 1990 Oct;9(10):3261-7, Wang *et al.*, *Brain Res Bull* 1998;45(4):421-5, Glencorse *et al.*, *J Neurochem* 1993 Dec;61(6):2294-302, Kofuji *et al.*, *J Neurochem* 1991 Feb;56(2):713-5. Shivers *et al.*, *Neuron* 1989 Sep;3(3):327-37, Wilcox *et al.*, *Proc Natl Acad Sci U S A* 1992 Jul 1;89(13):5857-61, Whiting *et al.*, *Proc Natl Acad Sci U S A* 1990 Dec;87(24):9966-70.

Transporter proteins, particularly members of the gamma-aminobutyric-acid receptor subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown transport proteins. The present invention advances the state of the art by providing previously unidentified human transport proteins.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human transporter peptides and proteins that are related to the gamma-aminobutyric-acid receptor subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate transporter activity in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in the brain.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule (SEQ ID NO:1) that encodes the transporter protein of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in the brain.

FIGURE 2 provides the predicted amino acid sequence (SEQ ID NO:2) of the transporter of the present invention. In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences (SEQ ID NO:3) that span the gene encoding the transporter protein of the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. 113 SNPs, including 32 indels, have been identified in the gene encoding the transporter protein provided by the present invention and are given in Figure 3.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a transporter protein or part of a transporter protein and are related to the gamma-aminobutyric-acid receptor subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human transporter peptides and proteins that are related to the gamma-aminobutyric-acid receptor subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these transporter peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the transporter of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known transporter proteins of the gamma-aminobutyric-acid receptor subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in the brain.. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known gamma-aminobutyric-acid receptor family or subfamily of transporter proteins.

Specific Embodiments

Peptide Molecules

10 The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the transporter family of proteins and are related to the gamma-aminobutyric-acid receptor subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the transporter peptides of the present invention, transporter peptides, or peptides/proteins of the present invention.

15 The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprising the amino acid sequences of the transporter peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

20 As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

25 In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

30

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the transporter peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated transporter peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in the brain. For example, a nucleic acid molecule encoding the transporter peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic

sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the transporter peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The transporter peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a transporter peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the transporter peptide. "Operatively linked" indicates that the transporter peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the transporter peptide.

In some uses, the fusion protein does not affect the activity of the transporter peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant transporter peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to

generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The

generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The

percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm.

(*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can

be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the transporter peptides of the present invention as well as being encoded by the same genetic locus as the transporter peptide provided herein. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 4 by ePCR, and confirmed with radiation hybrid mapping.

Allelic variants of a transporter peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by the same genetic locus as the transporter peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 4 by ePCR, and confirmed with radiation hybrid mapping.

As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a "-").

Paralogs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide, as being encoded by a gene from humans, and as having similar activity or function. Two

proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the transporter peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the transporter peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a transporter peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant transporter peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to transport ligand, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an

insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as transporter activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the transporter peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a transporter peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the transporter peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the transporter peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in transporter peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Accordingly, the transporter peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature transporter peptide is fused with another compound, such as a compound to increase the half-life of the transporter peptide (for example, polyethylene glycol), or in which the additional amino

acids are fused to the mature transporter peptide, such as a leader or secretory sequence or a sequence for purification of the mature transporter peptide or a pro-protein sequence.

Protein/Peptide Uses

5 The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a transporter-effector protein interaction or transporter-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

10 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

15 Substantial chemical and structural homology exists between the gamma-aminobutyric-acid receptor protein described herein and rat, mouse, bovine, and chicken GABAA receptor (see Figure 1). As discussed in the background, there GABAA
20 receptors are known in the art to be involved in medicating neuronal inhibition by GABA, and are the major inhibitory neurotransmitter receptors. Subsequently, they may play a role in human genetic predisposition to the development of alcoholism, as well as in the maintenance of psychomotor disease states, particularly schizophrenia.
25 Accordingly, the gamma-aminobutyric-acid receptor protein, and the encoding gene,
30 provided by the present invention is useful for treating, preventing, and/or diagnosing

alcoholism, psychomotor diseases such as schizophrenia and other disorders associated with neurotransmitter receptors.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, transporters isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the transporter. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain. A large percentage of pharmaceutical agents are being developed that modulate the activity of transporter proteins, particularly members of the gamma-aminobutyric-acid receptor subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates expression in the brain. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to transporters that are related to members of the gamma-aminobutyric-acid receptor subfamily. Such assays involve any of the known transporter functions or activities or properties useful for diagnosis and treatment of transporter-related conditions that are specific for the subfamily of transporters that the one of the present invention belongs to, particularly in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems ((Hodgson, Bio/technology, 1992, Sept 10(9);973-80). Cell-

based systems can be native, i.e., cells that normally express the transporter, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in the brain. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the transporter protein.

5 The polypeptides can be used to identify compounds that modulate transporter activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the transporter. Both the transporters of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the transporter. These compounds can be further screened against a functional transporter to determine the effect of the compound on the transporter activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the transporter to a desired degree.

10 Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the transporter protein and a molecule that normally interacts with the transporter protein, e.g. a substrate or a component of the signal pathway that the transporter protein normally interacts (for example, another transporter). Such assays typically include the steps of combining the transporter protein with a candidate compound under conditions that allow the transporter protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the transporter protein and the target, such as any of the associated effects of signal transduction such as changes in membrane potential, protein phosphorylation, cAMP turnover, and adenylyate cyclase activation, etc.

15 Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single

chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant transporters or appropriate fragments containing mutations that affect transporter function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) transporter activity. The assays typically involve an assay of events in the signal transduction pathway that indicate transporter activity. Thus, the transport of a ligand, change in cell membrane potential, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the transporter protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the transporter can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the transporter can be assayed. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain.

Binding and/or activating compounds can also be screened by using chimeric transporter proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a ligand-binding region can be used that interacts with a different ligand than that which is recognized by the native transporter. Accordingly, a different set

of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the transporter is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the transporter (e.g. binding partners and/or ligands). Thus, a compound is exposed to a transporter polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble transporter polypeptide is also added to the mixture. If the test compound interacts with the soluble transporter polypeptide, it decreases the amount of complex formed or activity from the transporter target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the transporter. Thus, the soluble polypeptide that competes with the target transporter region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the transporter protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S -labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of transporter-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies

reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a transporter-binding protein and a candidate compound are incubated in the transporter protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the transporter protein target molecule, or which are reactive with transporter protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the transporters of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of transporter protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the transporter pathway, by treating cells or tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in the brain. These methods of treatment include the steps of administering a modulator of transporter activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the transporter proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the transporter and are involved in transporter activity. Such transporter-binding proteins are also likely to be involved in the propagation of signals by the transporter proteins or transporter targets as, for example, downstream elements of a

transporter-mediated signaling pathway. Alternatively, such transporter-binding proteins are likely to be transporter inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a transporter protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a transporter-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the transporter protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a transporter-modulating agent, an antisense transporter nucleic acid molecule, a transporter-specific antibody, or a transporter-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The transporter proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in the brain. The method involves contacting a biological sample with

a compound capable of interacting with the transporter protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered transporter activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-

266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the transporter protein in which one or more of the transporter functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and transporter activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in the brain. Accordingly, methods for treatment include the use of the transporter protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still

considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the transporter proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or transporter/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to

the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in the brain. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates expression in the brain. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in the brain. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the transporter peptide to a binding partner such as a ligand or protein binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a transporter peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the transporter peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprise several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and

cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the transporter peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a “-”).

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or

at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 4 by ePCR, and confirmed with radiation hybrid mapping.

Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a “-“).

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. 113 SNPs, including 32 indels, have been identified in the gene encoding the transporter protein provided by the present invention and are given in Figure 3.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

5 The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

10 The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

15 The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

20 The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 4 by ePCR, and confirmed with radiation hybrid mapping.

 The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

 The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

25 The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

 The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

30 The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in transporter protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a transporter protein, such as by measuring a level of a transporter-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a transporter gene has been mutated. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate transporter nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the transporter gene, particularly biological and pathological processes that are mediated by the transporter in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in the brain. The method typically includes assaying the ability of the compound to modulate the expression of the transporter nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired transporter nucleic acid

expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the transporter nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for transporter nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the transporter protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of transporter gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of transporter mRNA in the presence of the candidate compound is compared to the level of expression of transporter mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate transporter nucleic acid expression in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for transporter nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the transporter nucleic acid expression in the cells and tissues that

express the protein. Experimental data as provided in Figure 1 indicates expression in the brain.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the transporter gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in transporter nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in transporter genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the transporter gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the transporter gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a transporter protein.

Individuals carrying mutations in the transporter gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a "-"). As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 4 by ePCR, and confirmed with radiation hybrid mapping. Genomic DNA can be analyzed directly or

can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a transporter gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant transporter gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a “-”).

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or

DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into transporter protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of transporter nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired transporter nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the transporter protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in transporter gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired transporter protein to treat the individual.

The invention also encompasses kits for detecting the presence of a transporter nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that transporter proteins of the present invention are expressed in the brain. Specifically, a virtual northern blot shows expression in the human and fetal brain. In addition, PCR-based tissue screening panel confirms expression in human fetal brain. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting transporter nucleic acid in a biological sample; means for determining the amount of transporter nucleic acid in the sample; and means for comparing the amount of transporter nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect transporter protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a

substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the transporter proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the transporter gene of the present invention. Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention. 113 SNP variants were found, including 32 indels (indicated by a “-“).

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising

one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified transporter gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector.

Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced

at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterotransporter. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to

express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as transporters, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with transporters, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In

addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

5 The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a transporter protein or peptide that can be further purified to produce desired amounts of transporter protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the transporter protein or transporter protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native transporter protein is useful for assaying compounds that stimulate or inhibit transporter protein function.

Host cells are also useful for identifying transporter protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant transporter protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native transporter protein.

Genetically engineered host cells can be further used to produce non-human
20 transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying
25 the function of a transporter protein and identifying and evaluating modulators of transporter protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the
30 oocyte to develop in a pseudopregnant female foster animal. Any of the transporter protein

nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the transporter protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, transporter protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* transporter protein function, including ligand interaction, the effect of specific mutant transporter proteins on transporter protein function and ligand interaction, and the effect of chimeric transporter proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more transporter protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<110> BRANDON, Rhonda et al.

<120> ISOLATED HUMAN TRANSPORTER PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
AND USES THEREOF

<130> CL001006-CIP

<150> 09/730,002

<151> 2000-12-06

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2819

<212> DNA

<213> Human

<400> 1

```

tgaggtgcac tgcctttcca cactctccct tctgtactca gccagctgct gctgaggtgg 60
gaggaaaagt cctggctggg agaattgagc tagtgcagca cacgtaaaaa agcgattccg 120
atgggtcctt tgaaagcttt tctcttctcc ccttttcttc tgcggagtca aagtagaggg 180
gtgaggttgg tcttcttggt actgaccctg catttgggaa actgtgttga taaggcagat 240
gatgaagatg atgaggattt aacggtgaac aaaacctggg tcttggcccc aaaaattcat 300
gaaggagata tcacacaaat tctgaattca ttgcttcaag gctatgacaa taaacttcgt 360
ccagatatag gagtgaggcc cacggtaatt gaaactgatg tttatgtaaa cagcattgga 420
ccagttgatc caattaatat ggaatataca atagatataa tttttgcccc aacctgggtt 480
gacagtcggt taaaattcaa tagtaccatg aaagtgccta tgcttaacag taatatgggt 540
ggaaaaattht ggaattcctga cactttcttc agaaactcaa gaaaatctga tgctcactgg 600
ataacaactc ctaatcgtct gcttcgaatt tggaatgatg gacgagttct gtatactcta 660
agattgacaa ttaatgcaga atgttatctt cagcttcata actttcccat ggatgaacat 720
tctgtccac tggaattttc aagctatgga taccctaaaa atgaaattga gtataagtgg 780
aaaaagccct ccgtagaagt ggctgatcct aaatactgga gattatatca gtttgcattht 840
gtagggttac ggaactcaac tgaaatcact cacacgatct ctggggatta tggtatcatg 900
acaatttttt ttgacctgag cagaagaatg ggatatttca ctattcagac ctacattcca 960
tgcattctga cagttgttct ttcttgggtg tcttttttga tcaataaaga tgcagtgcct 1020
gcaagaacat cgttgggtat cactacagtt ctgactatga caacctgag tacaattgcc 1080
aggaagtctt tacctaaggt ttcttatgtg actgcgatgg atctctttgt ttctgtttgt 1140
ttcatthttg tttttgcagc cttgatggaa tatggaacct tgcattattht taccagcaac 1200
caaaaaggaa agactgctac taaagacaga aagctaaaaa ataaagcctc gatgactcct 1260
ggtctccatc ctggatccac tctgattcca atgaataata tttctgtgcc gcaagaagat 1320
gattatgggt atcagtgttt ggagggcaaa gattgtgcca gcttcttctg ttgctttgaa 1380
gactgcagaa caggatcttt gaggaagga aggatacaca tacgcattgc caaaattgac 1440
tcttattcta gaatattttt cccaacogct ttgacctgt tcaacttgggt ttattgggtt 1500
ggctatcttt acttataaaa tctacttcat aagcaaaaaat caaaagaagt ctgactaaat 1560
tcagtagaat cttttgtact tcagtaactt gaagtttaaa tttaaaatgc agagagacca 1620
atggttaaaa tgtgaatagt attgtaacta ttttaaggcc ttcagaagta aataaagtag 1680
cagctttcag gctaattttac gtgaaactga ttagtgtcaa aatccagtag gttaaaatac 1740
tcacatattt ttacttaaat tttctttaat ttacttatat gttattataa ttttgaattt 1800
ttaagttcta tgattcatgt tttaaagatg gaatagtttt aatacatatt ttgtttaaat 1860
ataatctata attgttttgt aatgtaagac taattactaa tatttatgta gcaacttttg 1920

```

09/730,002
 2000-12-06
 CL001006-CIP


```

tgccgaaaaa agactgttaa tttgtttttt cttgcttttc atttgattac cctgcttgaa 1980
atacagtttag ttgataacat aaagccatag ttttcttgga ttttcttcca agatattgta 2040
ttcccaagaa gaaattgatt tttttttaaa ctatcagtta ctgaagactt atgaaaaggt 2100
caattttttac ctgtctttta atccagtcca ttttctgaca caatattaaa cagaacgcca 2160
gttgcatTTta ctttggtgat ttgcaaactt ggaatgaagc caccagtcac tttttaagaa 2220
tgcaagatga aaaaatcaag gtaaaatcta accattttat tctctgcttc atagcattta 2280
taatggtttg atgagggttaa cactgaaata ttaaaatctg caaatgcacc attaatggcg 2340
aattaaatgc cacagagaaa tggattatTTt tttgtcttta tggttttatg gaaagggtgtt 2400
ggcagcgccc ttacaattaa aagtatttga ggaaacaatt ggcttggaat ttaactggat 2460
gtgggttaat ataaattatt gaccctgggc aaaactgggg gttggggcgg ggtaacaatt 2520
gggaggaccg aaaattcttt aaatattgct gctttattag gcgctcgggg tttaaaacct 2580
tgataattga cgcccgaaaa aacagggggt taagggggccc cgccggcagg caaatttggc 2640
cctttttttc aatgtttcaa aggattcttc acggaccctt cgttcaatgt tttgcgaaca 2700
ctcgtttata ttttgccgtc gctctcagga ctgccgtcgt ccttacatta ttatcgctaa 2760
gaaaaacctg gggggggggc acagcgccgt ctcaaagaat ctgtgggttt actgagcgc 2819

```

<210> 2

<211> 465

<212> PRT

<213> Human

<400> 2

```

Met Gly Pro Leu Lys Ala Phe Leu Phe Ser Pro Phe Leu Leu Arg Ser
1          5          10          15
Gln Ser Arg Gly Val Arg Leu Val Phe Leu Leu Leu Thr Leu His Leu
20          25          30
Gly Asn Cys Val Asp Lys Ala Asp Asp Glu Asp Asp Glu Asp Leu Thr
35          40          45
Val Asn Lys Thr Trp Val Leu Ala Pro Lys Ile His Glu Gly Asp Ile
50          55          60
Thr Gln Ile Leu Asn Ser Leu Leu Gln Gly Tyr Asp Asn Lys Leu Arg
65          70          75          80
Pro Asp Ile Gly Val Arg Pro Thr Val Ile Glu Thr Asp Val Tyr Val
85          90          95
Asn Ser Ile Gly Pro Val Asp Pro Ile Asn Met Glu Tyr Thr Ile Asp
100         105         110
Ile Ile Phe Ala Gln Thr Trp Phe Asp Ser Arg Leu Lys Phe Asn Ser
115         120         125
Thr Met Lys Val Leu Met Leu Asn Ser Asn Met Val Gly Lys Ile Trp
130         135         140
Ile Pro Asp Thr Phe Phe Arg Asn Ser Arg Lys Ser Asp Ala His Trp
145         150         155         160
Ile Thr Thr Pro Asn Arg Leu Leu Arg Ile Trp Asn Asp Gly Arg Val
165         170         175
Leu Tyr Thr Leu Arg Leu Thr Ile Asn Ala Glu Cys Tyr Leu Gln Leu
180         185         190
His Asn Phe Pro Met Asp Glu His Ser Cys Pro Leu Glu Phe Ser Ser
195         200         205
Tyr Gly Tyr Pro Lys Asn Glu Ile Glu Tyr Lys Trp Lys Lys Pro Ser
210         215         220
Val Glu Val Ala Asp Pro Lys Tyr Trp Arg Leu Tyr Gln Phe Ala Phe
225         230         235         240
Val Gly Leu Arg Asn Ser Thr Glu Ile Thr His Thr Ile Ser Gly Asp
245         250         255
Tyr Val Ile Met Thr Ile Phe Phe Asp Leu Ser Arg Arg Met Gly Tyr
260         265         270
Phe Thr Ile Gln Thr Tyr Ile Pro Cys Ile Leu Thr Val Val Leu Ser

```

275 280 285
 Trp Val Ser Phe Trp Ile Asn Lys Asp Ala Val Pro Ala Arg Thr Ser
 290 295 300
 Leu Gly Ile Thr Thr Val Leu Thr Met Thr Thr Leu Ser Thr Ile Ala
 305 310 315 320
 Arg Lys Ser Leu Pro Lys Val Ser Tyr Val Thr Ala Met Asp Leu Phe
 325 330 335
 Val Ser Val Cys Phe Ile Phe Val Phe Ala Ala Leu Met Glu Tyr Gly
 340 345 350
 Thr Leu His Tyr Phe Thr Ser Asn Gln Lys Gly Lys Thr Ala Thr Lys
 355 360 365
 Asp Arg Lys Leu Lys Asn Lys Ala Ser Met Thr Pro Gly Leu His Pro
 370 375 380
 Gly Ser Thr Leu Ile Pro Met Asn Asn Ile Ser Val Pro Gln Glu Asp
 385 390 395 400
 Asp Tyr Gly Tyr Gln Cys Leu Glu Gly Lys Asp Cys Ala Ser Phe Phe
 405 410 415
 Cys Cys Phe Glu Asp Cys Arg Thr Gly Ser Trp Arg Glu Gly Arg Ile
 420 425 430
 His Ile Arg Ile Ala Lys Ile Asp Ser Tyr Ser Arg Ile Phe Phe Pro
 435 440 445
 Thr Ala Phe Ala Leu Phe Asn Leu Val Tyr Trp Val Gly Tyr Leu Tyr
 450 455 460
 Leu
 465

<210> 3
 <211> 82938
 <212> DNA
 <213> Human

<400> 3
 gatggctagg cagagaaagc gcctgcttcc accacctcca gcgtcttcgc ctgtacagtt 60
 gttccagctg ccagaggcct tctggttacc atggcaaccg tcgggctctg ctaggaactc 120
 aaagcagagc ctctgaagtt aggtaccaa cactgaggct taacagatgc aaaagcagat 180
 ggggagtgagg gcagagaact aggattgccc atctaataagg tctacagctc ctatggctat 240
 gacaacattt ataaaatcag acttctgttt ttgtgtttgt aagatgagat taagatgaga 300
 tagttcttgc ctggcattat tcattagtgt catctttcgg tgcccggcgg gagaagaaga 360
 ctgcatatcc tctctcacta ttaaggcggg ggaagggggg tggggggagc aaacgtggtg 420
 accaggaatg catggagcct gctcattcat tcattcattc attcaattat acattcaaca 480
 gagacaccgc cccaccaca aaattaacaa aatgattaac acaaattaac agaactcctca 600
 cgaacctggg gaaaaccaat taggcactgc ctgcttccca cgggtctctc ttggaagaca 660
 gatttagaat ccgaacgctg gagaggaaag aggcagtcgc cccgaaatag agtcagacaa 720
 gaggaagagc agaagggggg aaaggcagcc tctccctagg gagggattct gttcaaggcc 780
 actcaaacgt ttctccgccc ctgcgaggga gaagaggttt gtttttcttt tagaaaaagg 840
 gagccctttt ccctactggg ctcaatttct agtaggagaa cccgcctcc ttgggctact 900
 gattggctca actcttttta agataattta tgccaccatc ctctcggcct gattggctcc 960
 ctgggtcaat tctgctggga gtgcatacct acctgtttgg gaggtgcact gcctttccac 1020
 actctccctt ctgtactcag ccagctgctg ctgaggtggg aggaaaagtc ctggctggga 1080
 gaattgagct agtcgagcac acgtaaaaaa gcgattccga tgggtccttt gaaagctttt 1140
 ctcttctccc cttttcttct gcggagtcaa agtagagggg tgaggttggt cttcttggtta 1200
 ctgaccctgc atttgggaaa ctgggtgagta aattcagggg tttaaaacta ttctttgcta 1260
 tctacccctt tccccttatt cccctttctt ttaaaaatta ttccacttta agtaattaat 1320
 ttgcatttcc taacacatat gtggtattta ccacgtttt attttctgat tagatctatt 1380
 cattctcaga tagaaggaat gtaacatatt tctgaagcat gagattgggt cagttgtttt 1440

05613667 "033904"

tgacttcaag	cagagtataa	actagtgaac	tagtgaacta	atgaaggttg	agttactggg	1500
agggagtgag	ggggagtaga	aatgactttg	ccagtaaatt	tactgaatga	atgccaccga	1560
agaactttct	ggatgattct	gcattttaaa	aacaaagtta	tttcagaaac	ttgttggttac	1620
acagcatgat	atgtatgatg	tttttctacc	tactgctatt	gcttttcccc	tttcctcgtt	1680
tttggttaact	ttctggaaca	aaccagtggt	ttacatcaag	aaatattgta	aatctggcta	1740
aaaacaaatc	tatcagcatt	aaaaatgtat	agtatccaag	taaaaactga	ccgtgtatat	1800
ttaatttgaa	ccagctttga	gtgttttttt	ttaaaattgg	ctgtccatca	gtaattaaaa	1860
atgtgtggcc	acatgaatgt	cttggctctg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	1920
tgtgtgtgtg	tttgtgtgta	tggggacaga	gagagagaga	gagagaaaag	tgtcaccgat	1980
cacaaatttc	aaatttaaca	gaaacaaaac	tggactggaa	tacctttcag	taaaaatgtt	2040
taagtcattc	tggtagacat	ataaaatgct	caggtatcat	tcttcgatta	tacgtaggag	2100
tagtgggtga	cttgagctaa	aatacgttga	ttaagacata	gacacttata	aaggctctggg	2160
acatggtact	cataccatgc	aacttgtttc	aaataaatgt	acttttatca	ttacattatt	2220
aaaaaagtaa	ctctgtatag	aattcaaaaag	taaagtacaa	tcattatgaa	gatagaaaca	2280
tgcctttata	actatttgta	tcactgctaa	aatgaattat	tattataaac	gtaagtgtaa	2340
agctagagtg	aatttttttg	ccacaccagc	atatagaaca	agtatatttt	ctcttagaaa	2400
atagagattg	gattaaagta	ggctaagtaa	tattagacct	gttagaattt	tttgacagtt	2460
agtgtctttg	taggactgaa	gttttctctc	tctttttttt	taacctgtgt	cccttagatt	2520
cttaaatgtt	tgacacgaac	ataaacataa	tactggatga	attatttctt	gtatatattg	2580
tgtcgtaatg	gcttattctt	tgaatatatt	atagcatttt	ataaaagata	tttaggtata	2640
cacatacatc	ataatattac	aaaagtaaaa	acagattttac	agtagttttc	tgaaataaca	2700
accttttate	attttatacg	agtttttcctt	tcaaacaaat	aggcaagctg	cttagtaaaag	2760
agcatattat	ctatcccaag	agatgtttta	ttttaaataa	acatgttttg	ttttggggaa	2820
tattttgcaat	cattgtaatt	ggctccagag	ccaattacat	caagctccca	gaattgaaga	2880
gatatgtgct	gaaggacag	agtgcaggtt	tagcagcctg	gagacatgat	ctcatgtgtc	2940
aagggccttt	aacctctctg	aacacaggtt	catctatgaa	atgaaggata	ctagaccaca	3000
tatgatttct	taagagtttc	acattcaata	catctatgat	ttcaactgaa	acatgaaata	3060
agtcaatggg	tgcttcactg	ttttatctaa	gtacttttca	agagggaaat	tctaagtgtg	3120
aatactatgg	actttttaac	catatgtata	agtgaggaag	gtattgtaca	gttaacacat	3180
aatttttatt	gtcatagtaa	tgtagataac	ctggattatt	ttgtccagtg	aataaatata	3240
cattaaaaga	tatatattata	agctgttggt	ctaattagcc	cttacggcct	cagttatgac	3300
agttcatggt	atacatttaa	aagtaccccc	aaaaacataa	tttttaagaa	tgaggtagaa	3360
agctagcaat	aaaaggcaaa	tghtaagcatg	ctttctttta	gatgcatcat	atactgtttt	3420
cttcattgaa	tggaaaaaagt	ttgacacaga	aacacttttg	taggcaaacc	atgagaatgt	3480
tattagaaca	tgtttctttc	tccattgggt	tctatgagac	caagttggga	aatatagcaa	3540
ttaaaaggaa	ataaagcatg	cttttgctaa	aaaataaatt	taaaaaataa	tgttgcaatg	3600
aacatttctt	taaaataagt	aaaataaagt	agctacagtt	gggtgaactt	tttatccctt	3660
taatatattag	ttggaggtaa	cagatgctaa	catgtataaa	tcatgtacga	tcatatttct	3720
ctctcccttt	ccctttcttt	tggttagtgt	tcccttccaa	aactaaccat	aaatagttca	3780
tttgtaagca	ttttagtgtt	ttatatgatt	tccaaaatat	atcttgaaag	tcatctattt	3840
tggggacttc	ccaaattgga	agcatttttt	ttgtatttaa	tcagttgtaa	tgcttagttt	3900
cctcattaat	ttccatcaac	agtatagagg	agaaacacta	ttttattttt	tattgctgaa	3960
ctatgaaaac	ctatcctatc	attgtaacta	aaactgtagt	catggttcct	gtgtatcatt	4020
ctctaggaga	ctctttggca	tcttgacaag	ttcttttagaa	tctcttcagc	ttgatttaac	4080
ttctttgtaa	tgatgaaaat	aatctagtta	taatcaggtt	tttaaaaaga	atttaaaaag	4140
atatttttagc	tgttaccaaa	agtaacctac	aataagcctt	gactaagagt	tatacattga	4200
aagagctcca	ggaaaaaatc	acaattttct	tttttcagaa	caatttccat	gagtcaataa	4260
tcatactatg	tggcattcag	ttgcccgggt	gaaaatatct	aaataaactt	gcttttttaa	4320
atatattatt	ttcaagcaat	ttcagtagta	tgttttgtgt	gggtagtaat	gaatgaataa	4380
tttttttaagt	gtctagattt	ttggcttaaa	tttgaataaa	aatgttgatg	ataaccattc	4440
aaactttgaa	agcaaatgac	taagtaaaaga	tattccatga	tagttaaata	tgcaaaaagac	4500
tgagaggaaa	atgatttttg	aagctaaaaat	gcaaatgtcg	taattgataa	aacctctgtg	4560
tacagagtct	tttagtttat	ccaacaacatc	agtatctttc	ttccattcta	tactatgaag	4620
aacacttatg	gctaaggaac	aagaggaaat	gtttgatatt	tcattcaaat	attaaataga	4680
ttaattgaat	gtaagaatat	aaggatgaag	gtggaggccc	aaaggagggg	tgaaaagaaa	4740
gagtaaaaaa	aactgaaaaa	tgactctgga	gatactgaga	aaatgattac	caaaatatga	4800
taaatgactt	ccatgaaaaga	aaaggaagaa	tgagcttcaa	gagagtaatt	cacatatggg	4860

caggctttca	agtgagaatc	ctggccttta	ccaacagatc	taaattccag	tttcagctcc	4920
accattttctc	actggagtga	tttcagacaa	ctccttgagg	cagaaaatct	tcatttagtgt	4980
ggagttgtcg	tatgcatttc	cagaaatgaa	atttccaaga	aagaaaatct	gtgcaaagaa	5040
actagcactt	aaaattttat	atgaaaagaa	agtcataaca	gatattctgt	ggaattatag	5100
ttattctatt	atgatatagc	ctcgctaaag	tggttttaag	tgcagtagag	tgaaatatat	5160
ggatccttgt	gtgttcttaa	atgtctttta	agtcttttaa	atgtctttta	aatgcagctt	5220
atgcttattc	atggctaaga	cctagattta	tgacactaat	taaataaaaa	atatattttta	5280
atcaaaaatta	aaaatacacc	tctagaactc	taaagataaa	cagaggagcc	taaatttttg	5340
aagatatatac	gaaagaatgg	atcataatca	aggaagaaga	actaaagagc	caagagattt	5400
ctcatcagca	ataccagaag	actatgggga	aaatgtcttc	aaatctgaag	tcaagtattc	5460
ctcctagaag	attataactta	gccaaattac	ctaattgagaa	gaagaagaag	aatatagcct	5520
cgatattgtt	attctctcaa	ttttacagat	ggaaaaacag	gggaaaaaca	tttttgtgaa	5580
cttattaaag	gtcatcttgc	tgataggtgc	tgaaaccgaa	acttaaattcc	ctgtgttcta	5640
tacccttgcc	tgcattttta	actaccacat	ctctcagatt	atgaatgcaa	tataagtcaa	5700
aacaatgaac	attagtgttg	ataaattttt	taaaggaaat	ggagaaaaaa	agaggagaag	5760
agaagagaga	tactacagta	aattatgagt	tcattgataa	aatcacatta	agcagaaaaa	5820
tctactttgt	acctatcaca	cactttctgg	agtattgtgg	gtagtgccag	attattttta	5880
agaaaatttt	gacacactga	agtacttgat	gcagacaaaac	acacacacac	acacacacaa	5940
ctgtaactta	aaagcaaaat	tttttaattt	acataagttt	actgaggaat	tgaatggtta	6000
tottattttta	aaggatgcta	tatataatca	tgtttgaatt	tattatgtat	agatgtagta	6060
gacaaaccat	aaccaagtag	tagaaattat	atgaagggaag	attttctccc	aatacaaaag	6120
ttatctaaca	tggtaatgga	tagcttatca	ctagacatgt	ttaagcagaa	acttgttagt	6180
gcgttctcag	gctctgtagg	aggtagatta	aatttatagt	tttttttatt	ttcaaacatt	6240
tcattaagtc	tatcagtact	aaaataattt	agccagctga	agtcccatgt	ttcaagaaat	6300
agccttagct	attaattcoat	tcagtaaata	ctcactgagc	actaccgttc	attccacca	6360
tatttactca	gtgcccattg	tgtgttaagc	agggttctgg	aaactagata	taatggtaat	6420
aatggtaaat	aagacaaagt	cattgcccta	tagtgttcaa	ctagcataaa	taagacatat	6480
atttatatgct	agatgaacag	gatacagaaa	tcagaattcc	tgctcaatga	acataaactc	6540
tagtagtttg	tgtcaagata	gcataattaac	ataaacaagt	aaacacaata	aaatgctgta	6600
attgctactc	agaattctgg	aggaggtaca	gtgggagcaa	taaagaagtt	gccatcagtt	6660
gcaactctgg	gaaaatattg	tatgtttctca	gctgtcatag	gcataagttc	aggtcagtat	6720
gagggaaaga	tatatataca	tctagatagg	taggtagata	gatagataga	tagatagata	6780
gatagataga	tagatctata	tgtaatatat	atccttacat	ctgaggaaag	ggtaggctat	6840
atgttctgtt	agggggttga	ggataagggg	aaggcatatt	atactgtagc	tgtatatata	6900
tatatacaca	cacacacgta	tatatgtata	tatacacaca	tatgtatata	tatgtatata	6960
tatacacaca	catatgtata	catacacaca	tgtatacata	tgtatatatg	tatatagata	7020
cacatgtata	catatgtata	tatgtatata	gatacacatg	tatacatatg	tatatatgta	7080
tatagatata	catgtatata	tatgtatata	tgtatataga	tacacatgta	tacatatgta	7140
tatatgtata	tagatacaca	tgtatacata	tgtatatatg	tatatagata	cacatgtata	7200
catatgtata	tatgtatata	gatacacatg	tatacatatg	tatatatgta	tatagatata	7260
catgtatata	tatgtatata	tgtatatata	tgtatgtata	tatgtatatg	tatatacatg	7320
tatgtatatg	tgtatatgta	tatatgtata	tgtatatata	tgtatgtata	tatacagcta	7380
tatacatgtg	tatatatata	cagctatata	cgtatatata	tatatatgta	tatatgtata	7440
catatgtgta	tatatatata	tagagagaga	gagagagaga	caaagagaga	cagagagaga	7500
gagagagaga	gagagagaga	gacagagata	acattacat	cttgtgccaa	aggtagagtc	7560
tcacagtttg	gtagaatttt	ttgtatatag	atttccctata	tatttttaggt	gaacctttac	7620
tctggttgtg	taagatgtgg	cattaaaata	gaaacatcct	attgaaggct	atataattta	7680
gttttaatttt	gaaatcctga	aatgtagaaa	tcttttcttc	ccagaaacaa	tgtattctga	7740
gatttgacat	ttttccttat	ttgttagaca	aatattaagt	gactaaaaag	taatagagat	7800
tgtgctatgc	aatagagata	cagaagtaaa	cttaatatag	accctgctct	ccaacctata	7860
gctagaatat	aggatcagaa	agtttatgat	tataaaatag	tgagtttaat	atatattttta	7920
atagtcatgt	caagatgtga	tttggttctc	tagatggctt	taagactgct	agaatcaagg	7980
gttctagcca	ctaggtttaat	cttttaggcag	atttctagaa	caatcactaa	tggttgtaca	8040
ttaggtacag	tatgcttaag	tacagggata	tcattcacct	cttgatcatc	ctagatgggt	8100
ataattataa	tatttaatttt	ccagcacgtg	taatagagtg	tgttgagtaa	agaataacctg	8160
ttccagaaaa	gagccatatt	tactagcact	gagtagccct	gttggtgga	ttgaacttaa	8220
gtattaattc	agatgcctgt	attttataaa	ctaatttggt	accctctgca	ggtggcttat	8280

```

gaatgtatat ttgtctctgg aatttcttag taacaattca ttccaaaat catcagattg 8340
tcatgactag agtagactag actgttttgc tttttctatt cacttatttg tattctctta 8400
gtttttgaag tgaatagtct aagacaatct tttgtcaatt aaaatatttg acactcattt 8460
cagatcccaa ggcttagaaa aatatatatg tttctgcctt cagaggtagt aattccgatt 8520
tctttcttcc acttttaagg tcttttaata tgttttataa ttaataaagc tcaatataat 8580
gcctccatgt tgcaatatgt ttagcagcct tgaagaatgt atctgaagca tagatatccc 8640
aaagtctcag agagtgtgaa tgtcaccacg atttgtattt tttttctaag tctaacttat 8700
gttctgtcag gcactatgaa aatgaactta aaaatggaga actcaatgat aatttcaata 8760
tattgtctat tcagtgttca ttttttaaat agactataat aaaaattaca ttttctttct 8820
cttactccaa aaagcactca catagaaatg gtcacaaaat aacatttaca atttctgggc 8880
atttaccagc cttctgaaga tttcattgga tgggtggttt gtagttttct acactctgct 8940
ttgaggaaca aagttggcct gctaaacata gcaataaac aacaaagaca aaaatctaaa 9000
agttggcaag ctatactgtt ttattttagt tatttattct tttattaaaa atcaatcaaa 9060
aagctacaaa aatttcaaat ctataaccaa aggcattaag aaaaaaatct gttttataat 9120
tgtctttttt atcctctttc taaaaacatg ccaccgtacc aattaatttt ttaaattgat 9180
tctgatatta ctottttttc attagctatc ttatgtttaa tcttatattg attctaaata 9240
aataaactat gctctgatat gattggcaat tgtccgactc atttttgaaa gtaacagaga 9300
tatacagtgc taaaagcaat ttaaattgct atttgtttcc tattttagat gtaattttga 9360
agagatctat atcttcaatc ctaccaagga aaccatataa aatagactat aaatgaagcc 9420
acattttact tctttctcct aaatcatttt ttcacatctc aaataaaatc attgatgtat 9480
aatcttattc atgctctggg ctgtagacag aaaattaatt ttactttaca gtttgagttt 9540
tttactcaac tttatttttc tttacttgct ttttgaaaat attttcaatt gttgctgttg 9600
ttgaaacata agaaaaaaat gattttccct aatgccatgg tatatataac aatttactgt 9660
gtaagggtgg aatctaattt tgttttcttc taaaagagct aattagcatt tcaacacaa 9720
ttacggaata attcttgtat ttctaaactg ataattcact tttatcatat attgtatttt 9780
tgggatacat ctatatattt gggatcttat tttgaaagct taattttgtt ttcactattt 9840
attgctacaa tgaggtttta cttacagaaa ctttattttg ttttcttttt attcttgcca 9900
gaaaaagtcc atctaagaat tcccttttta aataatatct tataaattct catgttgatt 9960
gttctactta attttaatgg ttttaacctg ttacacacaa aaggtagagat tctgattaaa 10020
cctgtattaa atgcatcaaa aatttaggaa tcatacatat acaatatttt ataggcaaaa 10080
tatttacttt gtcttttttt ttctgtcctt atgttccctc ataaactcct tagttttcct 10140
cacgtaagac atttaccttc cctgttatag ttccgcttaa tcatttcata tgctatggtg 10200
tcattgtcaa atgagtaagt ttttttcttc tgtttttatt tttgatgaat ttgtattggt 10260
atcagaaagt tagtatgttt tttaaattag ctttatgtta gggtatctac cataattttt 10320
aaattaagga ccacagttat agtttttagt atgttcttga ttcatggtaa ttttatctga 10380
aaatagcaca ataggattta aatttttttt tctgttcact tacttattca ttgattttta 10440
ttaaattggg cacaaactat gtatcagatt tgggagtagg tgctggagg atcacattac 10500
atacccttgg ctctaaggta tttaaatacc taaatcctt caaattgctt aaagcttaga 10560
gacggtcata aataaataag taattaatta agatacagta tgattaactg catggtagag 10620
ttaaccacag gacagattgg aagcatacag gacacactta cctagtattc taggcttggg 10680
tctcttctgt catttcaagt tttcaaagca tctctctatg tatgttaagt gataggttcc 10740
actgtagttt cctttgtgct cataaaataa tcatagtatt gagtggatta tcaattctcc 10800
tctttaaaac tctggagggt tagtgcttgc ttcagctgca catataataa aattctgaaa 10860
gtttaaatct ataattgctgt tatttatcac tcttttactt attttgcttt tctatagctt 10920
ttgtacagtg aaaaattatt cagagctaca accctagaaa ggaacttttc taaaaaggg 10980
tttagtagct caaacatgaa ggcaaaaaag agagaaaatt tatttacttt ttgattttag 11040
gtctaaaatc agtgattttt cctacccttt gaaaagctag ttcttacagt tcttgacata 11100
gagtatatgg catcttctac tcagaagata cgtttctttt ctggaacttt tccctgaaat 11160
actgtgttta tgaaaatgtg tttgatgggt gtaaataaat actacctaac ttgtcatgaa 11220
caacagatca cagatagggc ctgtggagac aacatatatt tatttactaa agaattttca 11280
ttgaaaccct gtgtgtgaca gatgctgttc tagaaacttg tgatatgtcc atgaactaga 11340
gataaaacac cttctaccat tttggagttg gcactcagga gggagaaata cacagtaaac 11400
aataagcata aatatattaa attggaactt ataaaattat ctttttgctc aaatgggcaa 11460
atataattta atgggttaat actatttcat cctttatcat ttatgtatag gattttataa 11520
ttaaattaga gctatggtga tacgggttagg cttcatgtcc ccccgccaa atctcatctt 11580
gaattgtaat ccccataatc cccaagtgtc aagagagaga ccagggtgat gtaattgaat 11640
catgggggca gttaccctta tgcgtttctc atgatagtga gttctcagga gatttgatgg 11700

```

05313657-032804

ttttataagg	ggctcttccc	ccttctctca	gcacctttcc	ttcttggtgc	cttgtgagga	11760
aggtaccttg	cttccccttt	gccttcacca	tgattgtaag	tttactgaag	ccttcccagt	11820
tatactgaac	tgtgagtaa	ttaaacctct	tttctttata	aattactcag	tcttgggcag	11880
ttctttatag	cagtataaaa	atggactaat	acatagagaa	acaaataagc	cgggcaagga	11940
ggtgtggaat	ttggatagaa	gagcaaatag	tgatcagaca	ggcttcaatg	attaaataat	12000
atltgaacaa	atactataag	tgagtgaagc	tttcagtcac	gccaagttag	ggttggggag	12060
ataagaattt	gaagtgggaag	accataatgt	gggggacctg	tagttacttg	tagccgtgta	12120
gctatttttt	tttcaacttt	tatttttaaa	tcaagggatg	catgtgcaga	ttttttacct	12180
gtgtatatct	caggatgctg	agctttgaag	taggaataat	cccatcagcg	atatacgaat	12240
catagatccc	aatggtttagt	ttttcaaccc	ttacctcata	cctccttgcc	ctctctaata	12300
gtccccattg	ccttttatgt	tcatgagcac	ccaatgttag	gctcacactt	acaaatgaga	12360
acatggggta	tttgcttttt	tgttctgcaa	aaaacacaat	tttggtgttt	tttaatggct	12420
tttttatggc	tatgtagtgt	ttatgtacct	cattttcttt	atccaatcca	ccactaatgg	12480
gcacctatca	gttgattcta	tcactttgct	gttggtgaaca	gtgctgcatg	gaacatgtga	12540
gtgcatgtgt	ctttttggta	gaattatttg	ttttcttttg	gatatatacc	cagtaatggg	12600
attgctgggt	tgaatgatag	ttctgtttta	agttctttga	gaaatctcaa	aactgctttc	12660
cacattggca	gaactatttt	gcattctcat	caacagtgta	taagcatttc	cttttctcca	12720
cagcctagtc	tatatctggt	gctttttggc	ttttcaataa	tagtcattct	gattgggtga	12780
ggatatcttg	ttgtggtttt	gatttgcatt	tctctgacgt	ttagagggtg	ggaatatggt	12840
ccctatgttt	gctggccacc	tctatgtctt	cttttgaaaa	atgtctgttc	atgtattttc	12900
ctacttttag	tggggtttta	ttttgttggg	ttttcaattg	tttaagttcc	ctgtggattc	12960
tagatattag	atctttgtca	gatgcacagt	ttgcaaatat	tttctctcat	tatgtaagtt	13020
atctgtttac	tccttgacag	ttgtttttgc	tctgcagaag	ctcttcagtt	taataaggtc	13080
tcacatgtcc	atttttgttt	ttgttgcaat	ttcttttgag	gactgagcta	taattgcttt	13140
cctaaagctg	acatccagaa	tgggttttcc	taggttttct	tgtaggatta	ttacagtgtg	13200
aagttttaca	tttaaacctt	taatctatca	tgagtcaatt	tttgatgtg	gtgcaagtta	13260
agggteccat	ttcagtcttc	ttcgtatggc	tagccaacta	ttccagcacc	atttattgaa	13320
tagagagtcc	ttttctcatt	gcttattttt	gccaaatttg	ttgaagatca	gatgtctgta	13380
ggtgtgcagt	tttatttctg	ggctctcttc	tgttccattt	gcctatgtgt	ctgtttttgt	13440
accagtacca	tgctgttttg	gttactgcag	ccttttcata	tagtttaaaa	taagataatg	13500
tgatgattct	ggctttgttc	ttttttcttg	gagttgcttt	ggctatttgg	gtcctttttt	13560
tgttccatat	gaattttaga	atagtttttc	ctagttctat	ggaaaactct	tttggttagct	13620
taataggaat	agcattgaat	gtgtaaattg	ccttgctcca	tatggccttt	tttaataatat	13680
tgtttattcc	aatccatgaa	cgtgaaatgt	tttgcatctc	tttgagtcac	ccttggaattg	13740
tttttagcagt	gttttgtagt	tctgcttgta	gagatctttc	agctccttgg	ttacatgttt	13800
tcctaggtat	tatttttggt	gctattgtaa	atgggattac	attcttgatt	tgactcctag	13860
attgaacatt	attttcgtat	agaaattcta	ctgagtttta	aacattgatt	ttgtgtcctg	13920
aattacaatg	aagtggttta	tcagttccag	gagccttttg	gtggagttct	ttgggatttt	13980
cctgggtatg	aatcacacac	tttctaataa	gagatagttt	gatttcttct	ttttctgttc	14040
gatgcctttt	atctctttct	cctgccaatg	tgtctgtggt	agcatttcta	gtcttacgtt	14100
gaataggagt	ggtaggagtg	gccatccttg	tcctgtttca	gttatcaagg	gtgggtgcttc	14160
caccttttga	ccatttagtt	tgatgttggc	tgtgatttgt	catagatgac	tcttattatt	14220
ttgaggtatg	ttccttcaag	tcttagtttc	ttgaggggtt	ttatcatgaa	aggatgttgg	14280
attatatcta	aagctttttc	caogtctatt	gagatgatta	catgtttttt	tttaattatgc	14340
tcatatgatg	aatcacattt	attcattttg	atatggtgaa	ccaaccttat	atcccaggaa	14400
tagaacctac	ttgattgtga	tgaattaact	ttttgatgta	ctggtgggtt	cagtttgcct	14460
tattttatttt	gagggttttt	gcatctatgt	tcacaaagga	tatttgcttg	tagttttctt	14520
ttttcattgt	ctttgggtggg	ttttggtatt	aggggtggtg	tggcttttta	gaattagtta	14580
gagaggactt	aaaaggcaca	gagtggcagg	ttggatagat	taacaagacc	catccatctg	14640
ctgtcttcaa	gacacctgtc	tcagaggtag	taatgcccat	cagctcaaaa	taaagtttag	14700
ataaagtttt	aatgggcaaa	cagaaagcaa	aaagagcagg	gttcaactatg	ttatatcaga	14760
taaaatgaca	ttaaaccatc	aacagtaaaa	atggacaaaag	gagggcattg	cataattata	14820
aagggtgcaa	ttcaacaaga	agactcagtt	atcttaaat	tatgtgcacc	caacattgga	14880
gcacctacat	ttaaagatat	acatctagaa	ctacaaaaac	atttagacag	tcacccaatc	14940
atagttgaag	atlttcaacat	cccactgaaa	gcatttagcaa	gattgaagga	gaaaagtaaa	15000
gtaaaaaaga	aattctgtaa	ctaaatttga	cacttgacca	attggaacta	gcagatatct	15060
gtagaacact	cttcccatca	accagagaat	atacattttt	ctcatctgca	catggaacct	15120

actccaagat	caaccacatg	ctcagccata	aagcacatct	caacaaattc	aaaaaggtca	15180
aatcatacc	acctatagtc	ttggaccaca	gtggaataaa	aatagatata	aataccaaga	15240
agatctcttg	gaaccataca	attatatgga	aattaaacaa	cttgccccag	aatgcctttt	15300
gggtaaacaa	tgaagtaag	ggataaatca	aaaaattatt	tgaataaaat	gaagacagag	15360
acacaatata	ctaaagtctc	tgttatgctg	caaaagcagt	gttaagaaga	aagtttatat	15420
gctaactgtg	ggcctgaaaa	ggttaaaata	atctcaattt	aattatgtaa	catcgcacct	15480
ggggaaacta	gaaaaataag	aacaaactaa	ccacaaaact	agcagaagaa	aagaagtaac	15540
taatatcaga	acagaactga	atgtaattgga	gacacaaaaa	tttgtacaaa	gaatcagtga	15600
aaccttttct	aaaatgataa	acaagattga	tggactgaaa	gctagattaa	caaagaaaaa	15660
gagagagaga	agatccaaat	aagcacatc	agaaatggca	atggtgacat	tacaactaat	15720
cccacagaga	tataaaaagat	cctctgatac	tattatggat	acctttatgc	acacaaacta	15780
gaaaatctgg	agcaaagtca	taaattcctg	gaaacacaca	atcttccaag	attgaatcag	15840
taagaaatta	aatcctgaa	cagaccatca	ctgagtttca	aaattgaatc	agtattaaaa	15900
aaaaatacta	acaaaaaaaa	agccttgggc	cagatggatt	cacagccaaa	ttctactagg	15960
catacaaaga	aaacctggta	ccaattctac	tgaaggtatt	ctaaaaactt	gaggagacac	16020
tgtttatcct	aagagatatg	gaaaaccctc	caactaagat	ccaggggttt	gagcatagat	16080
atgatatgac	aagatctgta	tattaaaaac	atcactatgg	gtgttgtatt	gataatagac	16140
tatgttatca	atgcaagggc	agaattaaag	aaacccaaaag	gagacttttt	aaaataattt	16200
gagtcttcct	gatgaatttt	acctgtgata	aattcttctc	ttgccactta	ctaatagtaa	16260
gaacttttga	aagttatttt	aatctctcag	aaactgagtg	atcttcactt	ataaaaattag	16320
ggtaaaatgc	ctaccttata	aaatccacat	tcatgttttc	attcatacaa	caaataataa	16380
ttgagtgtca	acttagtgcc	catcacttct	ctagtcaacta	gagatgcac	cataaaca	16440
aaaaaaatcc	cacgctcatg	gagagtggta	aggatagtg	taaaggactc	cacactcgtg	16500
gatagtggta	aggatttaata	tgttcagaat	cctagcacaa	aacttctgtc	attagatcat	16560
tgtgagttag	tgaatttctt	tagtttgaa	caaattttaa	gattcttgag	aacagaaagt	16620
caacattttg	ctcctacttg	actgatgaat	gagtggtactc	tattaagcaa	agcaaagtgtg	16680
aattgaacaa	aatcaatttc	tagttcaagt	ggctgtgcaa	aacactactt	ctcacacgga	16740
gatttagtga	catataaact	gaatgtgtgt	gccaaaaatt	tactaagcag	aattttttca	16800
tatttggtct	gacctgaca	ggccataatg	cttgtttctc	ggcttaagaa	atgtgcatga	16860
aaagagtatc	aattcatgtt	cccttctagc	taaaaattcc	ccaatactat	atgtttctta	16920
gaattttttc	attccatact	tattaccatc	atttatacag	aatatttccc	aagtatttcc	16980
aattttactaa	tataataatt	tttctgtatt	tctaattgtc	cttctgacaa	atatttaaaa	17040
tgtgcagaat	gtgctaggaa	acctattgat	ttaccttccc	ttttttgtca	tatgtcctac	17100
tagcatagat	ttttgtctat	cacatataca	taaacatttt	accaaataca	taagtaaata	17160
cttattctat	tattgatcct	tctttttata	gtcgttatgt	caccacttgt	agaatagtat	17220
totgataata	gtctaatttt	aaaaataatc	atatgtataa	agtattttta	atgtcaaagc	17280
tttgagtaga	agtattgttc	agttggaaa	tttgtagaac	atgagattgt	tgtagagtaa	17340
caacaagtta	acagaatatt	tgaagaaaat	aaattaaaaat	aagataggat	ttttgatata	17400
tccaaaggca	atgtcaagag	aatcatcatt	tctattttaat	gattcaaaat	tatataattt	17460
agtatcacca	tcataattca	catatgagaa	aaaaagagag	ggagatagga	agcaataaga	17520
aaaagaagga	gggtgggaaa	tgaggtggga	ggaatgggtga	cagggaggaa	gaagaagaga	17580
aggaaaggaa	acaagaaaag	aaggaactga	gtaacaatga	ggatgggtga	catgtacatt	17640
tgtgtgtgta	gaatttagaa	tgactgcatt	aaaatatata	atagtaagat	gtataatcat	17700
gttaaaaaata	aagcatttgtg	tacattatta	taaggtatta	ccttaggaac	atatatatata	17760
ttcaattttat	atgatttcat	aaaacttttg	ctgataaatg	gcaaactata	atgagtttct	17820
agttaatgac	aaatgaaaag	aaaataatat	tcaagaggag	agaaaagatc	aatagacaat	17880
tgtaaaacca	cataatgggt	acaaaaaatg	gaatgaatga	ataaaaccta	ctacttcata	17940
gcacaatagg	gtaactgtag	tcaacataac	ttaattgtat	atttttaaaat	aacttgaaga	18000
atataattgg	attatttgta	actcaaagga	taaatgcttc	aggggatgaa	tacaccattc	18060
tccatgctgt	gcttatttca	catcgcatgc	ctgtatctaa	acatctcaca	ttccccctaa	18120
atatatacat	ctactttgta	cccacaattt	ttttaattaa	aaaaatgatt	aatgtagtgt	18180
gcgttttgag	ctggaaaagat	gaattaggtg	accatgcattg	acagattctt	cctcactcat	18240
gattcttttt	catatcaagc	ttaatgacct	agcaaaaaaa	aaattttacag	taacctccac	18300
atgtttttatt	tcatatacca	aaaaaatgac	tgaaaatgaa	ctgttttctaa	agtcaggata	18360
tgttacaagg	aaaggaggtg	ttaggtgaga	ggagttcata	tacctaatta	ttgaatttaa	18420
catattatcc	atcctgagct	taattcagtt	gagcagagta	gaaccaagca	ttcattgtgt	18480
tcaagctatg	agcgtttatg	tacatttgag	ttattaacag	tgctcttgct	gtattatgtt	18540

059413657-032004

ttgaggaagt	aaatctgctc	atttcctacc	aatgcaaaat	aaaacaatgg	caaaatgtct	18600
ttgcttttct	aactctttgt	caaagtcttt	tggtgtataa	gatttttttc	attttaaatt	18660
acaaaattgc	acttagacaa	accaaacaaa	ccaatgtaaa	tacataaatg	actaattatg	18720
gggtatcccc	tgggacccag	tgaaggactg	atcaacccaa	catcaggaaa	gctggggatg	18780
gggcagtctt	ggtgacagaa	attcctttct	tttctctagc	tatgccaaag	ggctgcacta	18840
caaattagct	attgctgcat	ttattgatta	aaattctcca	ctgtcaggaa	tgaaagtctg	18900
gctaaccctt	tggggctcag	gtgtctaccc	tattctaaac	aacttggcct	gagaaaggac	18960
ctgggagtat	aaacagacct	gcttaagcag	cctgcctggt	gattgctttc	ataccaggtc	19020
acagtgtagt	tcagtaggtg	gtcatactca	aatgtgccta	taaaactaga	catttcagga	19080
aggagagtaa	agagctactg	gaacctggag	aagaaaataa	ttctagccat	ctaattgggt	19140
gaaataggac	tccttcactt	gttttgcttt	ttgatgataa	aaaacaatca	atacaaagga	19200
aagtgcatac	accgactaac	ttgtagagaa	tatctatcat	ctatctatct	atctatctat	19260
ctatctatct	atctatctat	ctatctatca	tctatctacc	catccatcca	tctatctatc	19320
taactgctta	cctattttat	tatacaggta	agtaatttta	gaaaatttga	ggattatgtg	19380
tctggagata	gtgtgatgtt	gatccaaccc	atattggaca	ttcaaattaa	ttatgggtgt	19440
ataaatacag	ttacagggtg	atcatgttat	gtagtgtaat	aacaattttt	ctgaacttaa	19500
aaatttttcc	ctcaagtaaa	taattgtcct	tttaaaacat	cccttttggt	ttgcatgtac	19560
tagtaattta	aaataatttc	caaattttct	ccaaatctcc	taaatgtaca	ggcagccaag	19620
atatgaagtt	atagggctat	tcaaagtgtc	gggactgtct	tctaccacat	tacaaaaaaa	19680
cagtagtggt	tatatattgt	tgtatcattc	ttccatcagc	ctcagcaaac	tgacgtttgc	19740
ttcattttct	tttctaagcc	agtttggttt	tttcgttttc	tttttttctt	tttcacttta	19800
attttctttc	gcaagtcagt	tttagagacc	accacacctc	gccaatgggt	catctgtttt	19860
ttgatcttgt	atcaataaatt	tgcccaggcc	ttaattagca	agtatcattc	tgccagcttc	19920
cttctctttg	tttgtgcttt	ggatttccga	agtcctcctg	cttataaaaa	atggagtttg	19980
ggaatctggt	tttctatctc	cttgctactt	ttggttgatt	tctgaaaaag	gaaggatagc	20040
gtgccacttt	taccgttttc	aagagggaagt	tctgaattac	cttttataaa	tccagtaatg	20100
tttaatatat	attcattacc	taagggttta	cagaaatata	aaccttaaaa	gtttgatgag	20160
gaaactgtgg	ctcaaagatt	tgtaaaggatt	tgtacaaatc	ccacattacc	attaaaagat	20220
taataactaat	acttaaggct	tctatttttc	atgagctttc	ttctgggggc	tgcttttagtc	20280
atctttgcat	cccagcctc	cagaaatgct	ggatgctcag	cacacaataa	gcaccttaata	20340
gacatttgct	aaatgaagtt	taataaaaata	attgttataa	ttgtttaaag	agtgttaaaa	20400
tcccaactag	aatgtagttt	acataaggca	aaataaccba	tacatatata	taggtgtttt	20460
attttataat	aattacagga	atgtaatgca	atgaaaatta	gatactatct	catatccaat	20520
agattggaaa	aagggtattat	gtgactacac	gtaataatgt	attatggcta	tatgtaacgt	20580
tgtgtcattc	aaggatacat	aggtatatgc	taaaatgtaa	taactaacta	agattagtta	20640
atgtaacaca	ttagttatta	cattttacaac	aacacggcaa	taaattcatc	ttaacattta	20700
taatatctga	ggctggagaa	gaatgtagag	cgttttggtt	cttaaacgtg	gtattgggtt	20760
catagacagt	tgatgtttac	atcatttatct	acttttttaag	agtttctaaa	atacttcata	20820
acttcctttt	actacaatct	ctatgtatgt	tttgagttag	agtcattttc	ctaagtaact	20880
gaaatataat	gtggtatgat	cattgataca	tttgaatcat	cataatcatt	tcacatgtga	20940
atataataat	tatatgaaat	ttttggcact	aagtagcttg	gaggccttaa	tgaagtcat	21000
tatgtgttta	gaactccagt	ttcctttatt	tttaatatata	ggagttaggac	atgaagatat	21060
ctaaacgtgc	tcttacttca	aacactctat	gttcctagaa	atgcttagtg	taacgtatat	21120
agaacttttg	tagagattac	atgcagctaa	tcataacttg	aatttggctt	ctgattgcta	21180
tataaaaaata	tataaaatat	cagttttgat	aagggtacct	tactttttgt	ttcatttctaa	21240
ccacttatgc	tttttctcag	ctacctaaat	gagcactaat	agggatagca	aaaccaccca	21300
gtcaattgtg	tttattagta	tgaataata	ggagttagca	aatgacactc	aaaatttaact	21360
ctagtcaatg	acattcaaatt	gacaagaata	atagtggtag	tcatgatttc	tgaacatatt	21420
tattttatttt	ttaattataa	caggcaaaaa	tttggttttc	ttgattgagt	atatctacct	21480
attatcttga	tttatatata	attcactgat	taatttgcta	ttaaaatatt	gttagcccca	21540
cttttaaaaa	atagggtcat	gataacctca	aatttgtaaca	atatgaataa	ccaaatttta	21600
tattttatttt	aataaaacatg	gtggaacttt	tccctgtgtg	gaaataatga	ccaactaact	21660
aaccagaag	tttttatgag	gaagtaccag	tgacaaatat	ttattatatt	tactatccag	21720
tgaatttttt	ctgtacatat	actcagtaca	agtagcaatt	atcatttttt	tccaatgaag	21780
tatggattcc	aatcaatcaa	tttcacttgc	aattatgttt	ttttcacact	gtgggggaaa	21840
tagtctatca	ctatttcttc	agggtttact	cacattcctc	aggcacctta	ttaagctact	21900
ttgagtttca	tctctggttt	cttaacagat	taatggaaga	agctccctcc	cgtgtcagaa	21960

001006-03004

ccaggaaatt	tgcaataacc	tttatatgaa	ctaaatacac	ggtgactaga	tcattaatca	22020
tccaaagtag	gccattttgt	gatgagaagc	tgctatttaa	aatcattttg	gggcaacagg	22080
tgtaaaccac	gattttcctt	ggcaaaccct	aatgaatggt	taattttacg	atgaaaaata	22140
gtttatatta	attgactgac	ttgctatagg	catatatttt	tttctctagt	aataaggatg	22200
gattttcttc	ataattttat	ggcaaagcat	gccttgtctt	tacttaaatt	aacccttggg	22260
tattattttg	aaactcttgt	ttgttatgta	gaagagtaat	atagttatta	aagggttttt	22320
ccccacatac	agtaatttct	tgatcatcaa	gaaatttgta	tgatacatta	aaatagctga	22380
gtttagtaag	acaaaaagaa	ataataatca	tatactaaat	gactcaaaga	ttttctttag	22440
taaataagta	caggttgcaa	ttgaatgtga	gaagaagtct	ttgagaaaca	gatgtcaaga	22500
tgcccttaga	cttctaagag	agaaatgctt	gggaaacata	aatgagagga	ggagggagtg	22560
gcaagagagg	cttcaggtaa	caaagatggt	tgaaacgtat	gaaagtagag	gggggaagaa	22620
gtagtttggg	tcaaaaagaat	ctgagtgcaa	tacccttcta	agaaagtttt	ggccaggagg	22680
gtaggaatcc	tccaacacaa	attgcccttc	agaggggctc	ctagtctgac	aggaatggat	22740
ctgaattaat	attcctacta	tgttcagtta	ttgacagaaa	gcagcctgag	acagcttggg	22800
tctgggtgcg	actcagggct	catccttagt	actggcaggc	agctatgttc	tcacagcagg	22860
agaaccaagt	ggcaaatttt	cagggatgcc	acaataattt	tcttaagaga	gtatgatatt	22920
ctacctctct	ctaaaatact	ctatttagtt	attcatttta	taattagaaa	tttatttcag	22980
agcgttatct	ggtatataag	aagaagttca	ttcttaggta	ggcattcaca	tcgaattcac	23040
tccttttcat	ttattttacat	tatacaatgt	gtacatgaag	atttcaggat	actattaaca	23100
tgcaaaaaata	aattttctat	tgtctcacta	gtatagaatg	ttgagtgtat	acattattct	23160
gtttttccat	caagtccttt	aagctacttt	atacatagat	gacaattaaa	gcaaactgtg	23220
tttctgttat	ttttcatgta	atattaaatt	atgaaaaata	tttgatgttt	ttacagactc	23280
ttaacaagca	atagatgtgt	actttttggt	aataaatcat	tatttaaaca	tactcttatt	23340
gttaaaaaag	ttatataatc	ataggttaata	ccgtgataag	catcttaatg	tctgatgctt	23400
ttctgtgtca	ttggattgaa	aatgttttct	cctttattag	aaaatgagac	catctagatc	23460
aagattataa	tttcatttgt	ctttgtattc	ctagcactta	gcacattttt	agtacacggt	23520
attatctcta	tagatggaca	aaaacactgg	gtcagagatt	gagatcacag	tgcttttttt	23580
gtcatatagt	gtgcgactcc	ttgagtttat	gattctacct	tctctggagt	caagggctga	23640
gactacccta	gccacacaca	atccaaatgt	gtctgcta	tacgtaccgt	aggttttgtt	23700
gcggtcttgt	gaattaggaa	attcaaaacc	gaaccaccct	agattattca	ataaagggtat	23760
atgttgtgcc	aggaaaactt	taaggaagca	gaaaatatgt	cttcaaatta	aatcaacatt	23820
attttaaatt	ctgggtctat	tgtttaatat	ctgtgtgtcc	tggaattcta	taatctctta	23880
ttttctcatt	tttaaggtag	gtatattaat	ttttaataaa	ttctcagagt	attattttta	23940
aaaaatcata	cctgtttctt	ccaaataaat	atttatccaa	aacatgtaac	tgctacaaaa	24000
atctcccagc	ttccaaaaaa	catcaggaca	tgaaacacac	tataagtgt	atgtttttta	24060
aattatgaag	tactttttta	tcaccaagat	aacaccaatg	taatattcaa	agtaattaat	24120
gaacattctg	agcataactt	agatagctta	gagaaaagg	gtagataaat	tacctataat	24180
gaaatttctt	ttatgttatt	gocgaaactg	aggaaaagata	actttatagt	atttttgagg	24240
tcaattttca	tattacctct	ctactacctg	agaaaagggc	aatataaaact	agtatcacca	24300
tgtggatttt	agattatcaa	aatattgcc	ccataaaaaa	aaaaacagaa	aaagcttttt	24360
tttttttttt	ttctaaaatc	ttccatttga	ttgcaatcat	tattcaaggg	aaagataaac	24420
tgtttatatg	tgcaagtggc	tgccataccc	accagatact	tcattctata	aaactgaaat	24480
ttagaggtga	atgaaaacct	aaaatttaag	ttgttgccaa	tatctcaaca	agtctaatta	24540
ggaaataaaa	cattccagaa	gcacacacca	gcagagattc	ttaatacttc	agtatttttt	24600
tttactttta	ttttaagttc	aggggtacat	gtaaagggtt	gttaagtaaa	cttgtgtcat	24660
gggggcttgt	tttacagctt	atttcatctc	ccagggtatta	agcctagtct	ccattagtta	24720
tttttctga	tcctttccct	cctcccaccc	tcccactct	gatatgccgc	agtgtgtgtt	24780
gtcccccttc	catttgcgta	tgtgtttctc	tcatttagct	ccaacttata	aatgagaaca	24840
tgtgacaact	cagtatttct	aattcctcag	cttatatgta	cttattttact	tgagagaaa	24900
aataggaagt	tatttgaagg	aaatcatata	acaaaatttc	tatactgtac	acaagataag	24960
tagagaaaaa	taattactgt	aatgatgtat	gtaagtataa	atatgttaac	taatagtgtc	25020
tccaggaatc	tgcatagggt	tataagtga	actttgtata	ctttaaaaata	atgcatacct	25080
agaagtacag	gtcctaaagta	tttttactca	tatatataac	tatctcaaat	tggaacaata	25140
ttatagtgtg	ctcattgcat	actgcgggta	tcagagcccc	agttctacca	gagtagcagt	25200
gttacattgg	tcaagttaat	ccacttctca	cagagttttt	cttcatttgc	aaagtaaggg	25260
aatataataa	tagtagaagg	accgtgggtt	ttagggtctg	agaactggcc	tctgactact	25320
agaaagcttt	cttgagggtg	catacaaagc	acattctatg	tacttttagt	atatacttaa	25380

72

tactgtccat	acaacaatta	atgatatttt	tctattttct	ttttttcttc	tcctttgctt	28860
ttttattttt	aattatatat	tttctgcttt	gtcacacat	gtaatatgtg	taaatttggt	28920
taacaccatt	ttcccaatct	ttgcttttagt	ctttaggctc	tttcaatata	tgctggaggc	28980
tacaatcttt	aggaaactaa	cacaagaaca	caaaaccaa	taccacccat	tctcacttat	29040
aagcgggagc	taaagtatga	gaattcatga	acactaagaa	aggaataaca	gacactgggt	29100
ggggagtggt	aaggaggaga	ggagcaaaaa	agataactat	tgggtaatgg	gcttcataatc	29160
tgggtgagaa	aataatcttt	acaggaaaac	ccctgtgaaa	tgaattcact	tacgtaacaa	29220
accttcactt	gtaaccccg	acctaataa	aaagttaa	actaattaaa	aaaattctca	29280
tcatttggtat	attttgctaa	agttatttta	gtcatctctt	tattgagtag	atttttcaga	29340
aaagtgttgc	atgtaaaatt	atttttctat	tgatttgaaa	cttgaaggac	atagtggcta	29400
catagaaaat	cctcaattta	tgttttcttt	cactgacttt	tcactttaat	agcctaattg	29460
cactggtaac	ttgctttgtg	tgttattttt	gaaaagtcta	atgccagtct	attattcagt	29520
ccttgtaatt	tacttagtct	ttttttttcc	tcctggaagc	ctttaggatt	ttttaagaaa	29580
gggttcaact	gaagtctttt	cttagtatat	gtttcagagc	tggtcaattt	gggtagccgt	29640
ttttctagac	attgatacat	ttttttcaaa	atgtaatttc	aggttttctt	ttcaaccttg	29700
ggaaaatgca	tttacttctg	tatttatttta	tttaggaaca	taaagttgaa	tatttgttct	29760
gttacattgt	tggtggtttt	ttcttcacga	attccacta	taggtatggt	agtcacactt	29820
tgccctgcctt	gtatttgcac	aatattctct	ctgactttct	gacttctctt	catctcaggt	29880
gtatttctctt	cattattttt	ctgcctatta	ttaaactggt	tttagttgat	tcattgggtga	29940
ccatcataat	tttgtctttt	tcttctatatt	tttaaagttc	aaccaatttt	ttattttatt	30000
cttacttttt	gtgtatttgt	gtgtgcgtga	gaaatgtatt	tatcttgott	agtttctgaa	30060
ttatttttatt	caaggtgatt	ttttcttatt	cctaaatttt	tggttgaaata	tacttaactg	30120
tggggagtat	tttcttacag	tttcttctgc	ttattatttt	gtgtgggaat	tttccacaat	30180
tgatttgtat	tggtacttgc	tatcttaaat	ctgcaataat	tattttgtoc	ttttaatttt	30240
gtggtttctg	ggtgttctac	aagatcttag	cttattgatg	tattttaatg	tcacgatagc	30300
aaagttcaga	tactttggat	ttgattgtgg	tggtgggtg	tattatttat	tttgtggtcg	30360
gggagaaatt	ctgttacttc	caattcattg	attcttcttt	tccttgtaga	agcatgaggt	30420
actccccttg	acttttctt	gtcactatac	tatttctctaa	cctcacttct	gtttttgttt	30480
ttgacatttt	tcttgccca	gaaacaacct	attttgaaca	tcactctttg	acatcacaca	30540
tacttgtaga	tccttttct	cttagttctt	gtgctgggtt	ccaaatacat	gtttaaatac	30600
tttcatatta	atggtacatt	tactttcttc	agtggacttt	ggcatctacc	ttaggcccat	30660
cactgactcc	cttctgtctt	ctacagagcc	tatttcaggc	tgctgtgccc	accataaact	30720
gttgcccttag	agtagagatg	agacaatggt	tgaaagattt	ctttccagat	ttgggtatttg	30780
gggtttttatt	tgtttgcttg	tttgacattt	acagtcattt	tgaagtgtgg	gcattttttg	30840
tctttaagtt	atactgagaa	cttaattttt	gtttacttct	acttttctct	tttttttata	30900
tacattttttg	aaggaaatat	tgggagatga	agacacaggg	aatttatgta	aattttctca	30960
tgtactgctt	tagactttta	cgattttatt	ctttgtaatt	tcttctttgt	cacatgcagt	31020
agttagtga	tattgtttaa	tttcaaaaa	tagcctgtgt	tctaggtata	tttttggtat	31080
tggtagctac	atttaatttaa	tctgggtctg	aaaacataaa	cttaagtcag	atttataatt	31140
cttggtgctc	aaatgtatag	atgtataatc	ctatacatct	actgagagag	gtgtgttaaa	31200
aatctctgac	tacaatgatt	gacttagcct	ttaaaaattt	ttagtccagt	aagattttgc	31260
tttatacata	ttgaaactgt	tttaggtaca	cacaaatttt	ggattgtttt	gtctcctttt	31320
ttaaaaaaaag	ttttactttt	atttttatgg	atacatagtg	gatgtatata	tttatggggg	31380
acatgaggta	tttttatata	cacatagaat	gtgtaataat	cacatcagga	taaatggggg	31440
agtcacctcg	agtatttatc	atttcttttt	ggtatgaaca	ttccagttat	actcttttgg	31500
ttattttttaa	atgtatgata	tattattatt	gactgtagtc	accctgctat	gctatcaaat	31560
aatagatatt	atttattgta	tctgactata	tttgcataca	cattaaccat	ccccactccc	31620
ctacctcact	accctctggt	aaccatcatt	ctatttcta	tctatgagtt	caattgtttt	31680
gattttcaac	ccctacaaat	gagtgagAAC	atgtgacttt	tgtttttctg	tgccctgggtt	31740
attttcactta	acataatgtc	atcagttcta	tccatgttgt	tgcaaatgac	agaatcttat	31800
ccttttttttt	gtggatgaat	ggtactcaat	tgtatacata	taccaaattt	tctttgcat	31860
gggttacacct	tctgttgaat	agcagtttta	tcaattaggaa	acaatctttc	tgggtctgtag	31920
aatacatttt	gctataacat	ttaaatttgat	tgatattata	tatctatatg	tgttttcttt	31980
gagttatgat	ttgacttttt	tttttttttt	tttttaagat	ggagttttgc	tggtgttgcc	32040
caggctggag	tacaatgggtg	caatcttggc	tcaatgcaac	ctcctccttc	tgggttcaag	32100
caatttctct	gtctcagcct	cctgagtagc	tgggattata	ggcatgcacc	accatgtccg	32160
gctaattttt	tatttttagt	agagatggag	tttctccatg	ttggtcaggc	tgggtctgctg	32220

74

ttctagcgat	ggggacatgg	gggagatata	aagaagattg	gtctacgttt	taactgggca	35700
tagttgcaac	cttagtccttg	ttactttgtta	cttctttttg	gggtaacatc	aagtggcctt	35760
aggtgtaagt	taaaggcttg	acgtataaat	aatgactttt	atgggttagga	ggtataaata	35820
attttttggt	gtttttcata	ggcattttat	atggctcaat	tccctggctt	gggaaatctg	35880
gatgctcttc	tgtatcttcc	acctgcctta	actcctttca	tcagcaatct	agtctgaaga	35940
tatgtgctgc	tgatgcacaa	ctctcttgag	gtcaaaaaac	ctgtgagacg	ctgctaacca	36000
agttattttc	ttttggcgtg	ctttataccc	acaagaaaat	aaaatatcct	aaaagtgcta	36060
aagagtaggt	gtgcagctct	tccctctgctt	tttcttcttt	aagtcctatc	acttctattt	36120
ttccttttcc	ctgatcaggt	atgggagcaa	gagcaggttg	ccaagtctgc	tgcctaaaca	36180
tatatttaag	tagagaatga	gaaaccagcc	ttcaattttc	ctgagaagac	attcctggga	36240
atacacacac	acacacacac	acacacacac	acacacacaa	acacacacaa	acacacacac	36300
acatgccatt	taaatataga	tatgtggggt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtattt	36360
tatagtgggg	atatacacta	taccatgcct	gtattgtaag	gtgcgaaaag	taattttataa	36420
tgaaatattt	ggcattctat	atgcgaatgc	tcaggtttta	ctacactgaa	agaaattatg	36480
aagcttccag	atgcttgcat	ttatacacag	aatgactgca	tattggaaag	cctgcgcaag	36540
ttaagagata	caaatgtgtc	ttatcagtga	ctccagta	atgagcaaca	ttgtgattgg	36600
tgatatggct	atctaaaatg	gtgtacaaca	ttttgtaaaa	ttcagaacaa	aataaatacc	36660
atatttctgt	gttttatattt	tcaggaattc	ctagcaaaaca	cagaaatact	ttgtgtttat	36720
gtgtaaaaaa	acttatcttc	taggttcaga	ctataaggca	tactgtttta	cctatatgaa	36780
tgctagacag	gatattcaaaa	atcttggttaa	atgtgggaca	atttttctatt	acatatgact	36840
gtttttatgaa	gttcaagcac	ttcatccact	aaacactact	actaacaatc	ccaatttttg	36900
tgacgaacac	aaacagctgc	ccagaaattt	tcaaaataat	ttacaggcga	catgggtgttc	36960
tcattgagaa	ccactgggtga	acattgttca	ttaaaactta	ctgaaatgat	agaaatgttg	37020
ttctgtattg	accattttttg	gaagcactgg	ccacaagtgg	ctgttctgta	cttgaaatgtg	37080
gattagtggt	gttaagaaat	atttttatatt	cattttaaatg	caatttaattt	ctattttaaat	37140
ttaaacagcc	atatgtttct	actggctaca	gtattagaca	gctcagcctt	aagaatctac	37200
ttatccctat	gtaccttttc	atccacatat	atctaaatgt	aacctggaaa	ggttgattgt	37260
caagtgggtg	ttgacatttc	cttggttaatt	aaggtctcac	aagaagtatc	taaactcgag	37320
cttggtatct	atctttgggt	caaagacttc	tgctgtatag	tctagggaaa	ctggaacatg	37380
ttaatcataa	ttatttcaaaa	gtctatatgt	ggtatggacc	tggagttttc	cttcttgagg	37440
catgtttata	agctctacca	tttaataact	ataggttttc	tgtaactag	atcgatatatt	37500
tatttttctc	aattttataa	cttgtaacat	tattttcatt	tttatttact	attgaaaaat	37560
ttatctcctc	ctgacttttt	cctgattcat	ggattatttt	ttaaattgta	tttaatttct	37620
aaatatttag	ggatttttaa	agatatcttt	tttattaatg	acttctattt	taactttgtt	37680
gtggtcctag	aataacttta	tattattttca	accttatata	agcatgacac	aataatgaaa	37740
accaggaaga	taatatgtag	atattactac	cagtgaatac	ttggccccac	ttcccaattg	37800
tctcaataag	gtattcacag	aaggatgatt	cagttcagaa	ataatattac	atttttgttg	37860
ctttactagc	gaatttcacc	acatatagaa	gaaagaacta	ctataagtct	tacacaaaat	37920
ctttaaaaaa	aaatagaaaa	ggatactcaa	ctcatttcac	taaaccagca	aaatcctgat	37980
aactaaacca	gacagggata	ttacaagaaa	aaaaaattat	acgctcatat	ctctcatgaa	38040
tgtatcggtt	gaataaaaaa	gttggtacat	ttatgagaaa	caattatttg	caggtaaaaa	38100
cagttgactc	cctgaataaa	ctttcagaaa	atcatcatag	aggctgctct	gcctatggag	38160
tagccattct	tttattcctt	tgctttctaa	aaacaaaaca	aaacaaaac	ttgctttcac	38220
aataaaaagaa	aatattcata	gaaatttgag	taggttgata	aaaataaacc	tgtatatggg	38280
caaggaccag	aagagagtcc	aaaagcaatc	agttgtgtaa	gagtattcac	actaggtggg	38340
tgcttgcttt	cttggaattg	tacctatatc	atttacaaac	aataatatac	ttggtgaaat	38400
gaaatattta	aatttgaaat	aaaattttata	taacataaaa	taaagaaaaa	aaattatgtg	38460
tgataactgc	aatatctgaa	taagtgggac	tatttctctt	ttttaaggcc	tattctgtga	38520
caataagaaa	agtctataaa	actctccagg	attagagttt	ttaattgaac	taaattaatt	38580
ttaatataca	aaagtgacca	aaaagctaca	taatctgcta	aagatatctt	agtgaatta	38640
gaatagtgtg	tccaagaaag	catgggttaa	atcaattata	aagaaaagat	ttttaaaatc	38700
tattgtagtt	tatattttct	cactgatata	ccaattacc	acctctacaa	ttggctgtcc	38760
taactattgc	ctttgctact	tggaaatctga	atgtggaaac	tttattgatc	acacagcttt	38820
tttcttttta	aaatttttaaa	cttttcttcc	tatttctgtt	ttctggatgt	tctttaaaat	38880
gcccgaacc	tctaattccca	ctggggagaat	atctacttac	atcctgttga	gcaacttaaa	38940
gcaaagtact	taaaataaca	gcacagaaag	aaaccttgca	aaaataatcc	agtgctcatct	39000
ctacctgcct	tcacattttat	atatacatat	gtgatttgcc	atggcagtac	atgggtttct	39060

catgcattgt	tatgctattg	caaaaaagag	atctttgatc	tccatgtgta	ttgatgaatt	39120
aaacatagtc	tgtgcttact	ttcaagatgt	tatttagagt	ctggaatgac	aattttctgt	39180
tagttttatt	tttctagttt	taaattaaaa	tataatgtct	tagtttttga	acatcactca	39240
tatcttgcaa	ttgaataaac	tttggattga	gaaaatcata	gtgctaagaa	agagtactta	39300
gtcaatttcc	aaaagcctgt	attttaagta	atattactca	ttagtattgt	tcatagtatg	39360
aatcaaagca	acaatgagtc	attattaaga	agcaattaca	gattttctgg	tttttttctt	39420
aattatat	taaagcatat	caccagccgt	ggtttcaaaa	ttaatatgaa	atatttttagt	39480
ctacttcctg	gtttattcat	ctttgtcgg	tgctgctgca	gattcaagga	tcagatcaaa	39540
tagcatctca	agtttacaaa	atgggtgaaca	gaataactca	aaaagattcg	tcatagcttt	39600
cacatacata	tttaacaatt	gttaactaag	tatcttctat	attagtagca	atattggact	39660
tactgatatg	attattcttt	tgtctcattc	tgggtgcaat	tgctctatgc	tcatcataaa	39720
acacttcagt	atctctgctt	ctatgttaact	catgctaata	tgccagagtt	atttttctaa	39780
tgtatatgtt	tatcttataa	ttggctaaga	aatgttgatg	tcaggaaacc	aagacagtta	39840
actcatatct	ataacctctg	acactgtaaa	tgtagaaaac	atztatcctc	taaagaaaat	39900
tgtcccagga	acaatttccc	tctcaatgta	gacaacgttc	agctttgaat	atgtgaaata	39960
aaagtaggca	ctcggcggtt	gctgatttat	cagtttaact	gaaaaaaaaa	atctcttatt	40020
atataagaat	ggtatatgtt	ttataaccagt	tataactact	tataagtgg	aataaatgta	40080
cttacagatg	gaactaataa	atctattgtc	tcgttggtta	ttgactttca	tgaaaatcat	40140
tctgtaaatc	taactataac	tggcttaatt	ttatatattt	tggctacaat	tcttagctag	40200
tcagaaacaa	aggagaaaca	gttaaagcaa	ctcatctcag	aagtcatcag	cctatgttaa	40260
attgtaaaac	cttctataaa	tcatatgacc	tacttttatt	taagatatct	aattgatggc	40320
tagcaaattg	aattgcattt	tcaagtattg	tgtaggtcta	gaaacatagt	tttgtgtact	40380
aaaattaaat	ctgtgttatt	aaatacgtat	gcaatataat	aaaatttttg	cttgaggaaa	40440
aacaaactat	gtaagtgttg	tgtgttttat	taagtgtatt	tttaattgca	aggggtaaaa	40500
tattgttcta	aaatttgata	gacattgttt	tacaaaatgc	ttcaagtttt	tcactgaagt	40560
ttggtagctt	gagggctctg	tcagattgaa	aaatgccaa	accaggagac	tgagactgaa	40620
tagatattca	gcctgccaga	cttgtgtgtg	tattaaaaga	ataaatacat	cttgcctggt	40680
caatcatagc	attgcgatta	tgagtataa	gtactcggtc	atataggaaa	aacagatgaa	40740
gtttgaaaag	gataactgag	catgtatgac	taaaaacatg	ccacaataac	agaaactttc	40800
agtctacaag	gtcatttttc	tgcattatct	catgtgcttt	atctctatat	aattatctgt	40860
gactgatttg	caaaccaaaa	agattgtttt	ttataaaggc	tacatttctg	catttcataa	40920
taaaagatgt	tttatgattg	aagtgttaat	atataagtag	ttattaagat	ttctttaaaa	40980
ctaagattta	tttaaaacaa	tgtctaatac	tatctttgac	ctttcttttt	gttgatattt	41040
tgcagtggag	cccacggtaa	ttgaaactga	tgtttatgta	aacagcattg	gaccagttga	41100
tccaattaat	atggtaagta	agaaatatta	aaaataataa	ctgccacaga	gatctctcgt	41160
gtatacctcc	aaaatagaat	tcgcatgtga	gtaatttttt	atagttaatt	ctgaatgaat	41220
tcagatcaac	atgtacaatt	actttgtgcc	aggatttggc	aactggagct	aaactatacc	41280
tgtatctaag	aatctgtatc	ttgttattcc	aatgtgctgt	gtttggtaga	actatgcaac	41340
ggcataacaa	attatcattc	gatccacca	gaagaggtcc	ctgaagaaat	ctggaaagaa	41400
aaaaaatcag	gaagtctgag	gggaaagtag	tcaatcagag	aaaaaaaaaga	cttactggca	41460
gagaaagcag	catgcacaaa	gtcacagagg	aatgaaacag	cctgataaga	tcagggggac	41520
ttccagcaat	ttgatgttac	tggaaataaa	agtgtaaagt	aaggaattga	ggcaaaaggg	41580
atattttaagg	atcttatcaa	ggacagtctt	gtaagccatc	ataataggaa	gcttgtaa	41640
tgtattatgt	ataaattgta	tccttgggga	gaagataaat	gacaggaaac	attttaccag	41700
gggagtgtca	aatgcagatg	tacattttta	aatacatctc	tcttattgca	gtatagaaga	41760
tggaggtaag	agtcacaagt	agggttacag	agaccatta	caggttattg	tgttagtctg	41820
gacaatagat	cttcaagggc	ttacctactg	cagttgcaat	caagctggag	atgaaaataa	41880
tacatttttt	tcaaaatagg	taattttttg	agaatttgg	gacttattaa	ttatgtgtaa	41940
cctgactcca	aggtattgaa	atgaagagac	tttgaagatg	gtagagtc	tatcagatgt	42000
aggaatacag	tgcaggccta	tgtaccccaa	atgctcatgg	tgcattcagc	aggaatggtc	42060
cagcagcatt	ttgccttata	gatttgctat	caggaagaa	gttaggacaa	aatatagatg	42120
tagatgagaa	agagaaagga	gacaagcagt	atataataaa	gctcaagaca	ttaagaatgt	42180
gtgagatgat	ctaagagaag	attcagatta	agaagatcac	agatctgagg	accagatggt	42240
gtaatgagat	gaatgtatgt	aaaatatagt	tgcaagaaag	agaagcacag	aaaggggaca	42300
tattcagaaa	agtcagaagg	gtaggagcaa	aaccgagaga	agatgatatt	ataaaatctg	42360
agaaagtaaa	gaggataaag	aaggagggaa	tgataatgag	ttacaaatgc	acatgagaat	42420
acaataaaat	ggattagata	gtgataactg	gcattcacaa	gtagggtggtg	ttgattagct	42480

09313657.032601

ttgttaggag	aaattcagtt	ggaatggtga	aggcaggaaa	gggataacag	ctggattgca	42540
tatccctcct	tcatcttcaa	aattgatctc	atgtccaaac	tttttgccat	ggatccctga	42600
gaataaatat	ggttcctgca	gggatagtg	tgtagtcatg	tgctggtttc	tgggtgctag	42660
ggagtgagtg	ggaatgtaga	tttaatttta	taaactacta	ttttgagaaa	catggcttag	42720
aagaagatga	gataagagta	tgggagtaga	gaaggtagag	gagagaagaa	gagaaaataag	42780
ggagggagag	ataatttgag	aaaatatcaa	gaaaagacca	tttagaatga	aaactgagaa	42840
agtgtgatat	tcatttttta	tgttttatac	atagaataaa	tagtgtttta	ttactaatga	42900
agagcacaat	taacaatgcc	aatatgagca	accagcatta	ttttaaaaaa	gacttcataa	42960
tgttcagttc	caaacgctaa	taatcccatt	tctgtgagtt	caaactactg	gatcccatgg	43020
aatgcattca	ataatgtctt	ctactcctct	ccttcgcttt	tgcccttggt	atttgttcac	43080
taggggttgc	acaacgaagt	agcacagact	ggtggtttta	gcaacaacaa	tctatttctt	43140
gcagttatag	aggccgaagt	tccaagatct	atgtcattgg	cttgcagaca	gctgctttct	43200
cactgtgtcc	ttacatggtc	ttttctctgt	gcctgtgcat	ccctgatata	tccttttgta	43260
tccaacgttt	ctcttcttat	cacaagtcat	attgtattag	ggcccaccct	aaaagccgca	43320
tgttgacttt	aaatgctctg	cctttaaata	tagtcacatt	ctgaagttct	cgggattaag	43380
tcttcaacat	atgaatttag	ggggttcagc	ccataagact	tccaagata	attctcaaat	43440
atcttgctga	ttttcctctt	cattttatatt	attcaatatt	ataatagtca	aatcatccct	43500
ttattacagg	ctttaccact	ttatagctgg	tgaaccttta	taatatgcct	aacagttgcc	43560
tatgtctgct	tactcctttg	gtgtgacagt	cacactcagt	ttttctccta	ctcaatgtaa	43620
tcttcacatg	actctcatta	acttttacag	ctgtctgcac	tgattactgt	agtgatagca	43680
ttctagccat	gccataagtg	acatctttgt	cacattagtg	caagaaaata	tgaactggct	43740
gtggataaat	tgaatttgag	aggatgaagg	aatgttccaa	tagactgacc	tagaagttag	43800
tcttaaatta	ggggctaaga	ttcagggaaa	agttgaggct	ggggccaaat	ctttgtcagg	43860
aatttgctaa	accctctgtg	gatgggggaa	aacagcctta	aagtgaagag	agaagaacaa	43920
agaaatgaat	agcagacctt	aggggaacaca	tgtattttaag	gacctgaaag	atgaagctga	43980
gatagtaaat	gaaacatcaa	ggttgccaag	aaagtgacaa	aacaatgacg	tggaaaccaa	44040
gggaagagaa	agtttcaaca	aatcttttta	gtattagttc	tactcaaaca	aattactata	44100
caataagctg	atttttgcaa	acagaaaact	tttaatatata	acgtgtaata	ggtgcatttt	44160
taacaatatg	caaatatcag	atttcataaa	gtagtacatt	gctaaaatgt	tatatcttct	44220
ctttagtgg	gtattgtaat	gttttagtttc	taagtaacca	atgcctagag	aaaagaagcc	44280
tgtttgagca	tatgggaatg	ttggcatttt	catgaattgt	atatttttctg	tttgggggtt	44340
atgttttcaa	tttgcatgaa	tacattaaat	aaaaatatgg	tcattaaagt	aaagagaaac	44400
aagattcaag	gatgtattgg	tattctagaa	acactttgac	aaaaaggcta	tcacagactt	44460
ttacaatgat	cttcttatat	aaaagattaa	tggtatttaa	gatttagttt	ccccagtgga	44520
gcaaattaca	cacttgtaac	cttcaacagt	gattgtttta	aaaaacctaa	ttaaataaaa	44580
ttgctattga	aactagtgtg	actaaaacag	gtgaaaataa	acaaataaaa	gaaccccatg	44640
gaacacattg	taatgtttat	taaaaagaaa	agcattgtga	tactagggat	gaataagaaa	44700
taggttaagt	agaagactat	tttaactaaa	caatacaggg	aaattagaat	tcacttttca	44760
taaaatgagc	aaaattgctt	tgaatgtgag	tgtaatcaag	gaagcagatt	tgagcatatt	44820
taagcaagag	ggtgccaccc	ccttaagtac	tcatataatt	taagtgacca	gtaagaaaca	44880
aaattaatat	attagttttt	tttaagttag	ccaaaatttg	ttttgctgca	aaatgggtgt	44940
ttctttttta	acttggtatc	ataaaatgca	tttccataga	tttctgaata	atgcctaaaa	45000
cgacattatt	ctacaatgga	atatgctttt	cttttttacag	tcaaaatgag	taacagatgc	45060
actaatgaag	actggggaaa	aatagagtaa	ccttgaaaga	agctgttcaa	ttttatatat	45120
ggtttgatta	taactatctt	caatataagc	agaagaaaaa	tatgcaatca	gaattgaagt	45180
caagaacatg	ttactagagt	ttatgagaca	aatggtttat	aacctcgtaa	cttttcagat	45240
ttgggggata	ttatgtatct	ttactcttct	gagtaaatat	ctcacgggtg	tgggtaccct	45300
ctgactggca	agtacaacac	tcaagggcc	tagcaattct	ttaaagccaa	gtcaaaattg	45360
gtaaaagctc	caggaaattc	aagatgatat	tatgggtatc	taacaggagg	ctacagtata	45420
agattagtta	gtacaaaagt	agctaccact	aaaaattcca	ggtataattt	ggacatgaat	45480
ttaattgtag	gaactgggat	acccttatat	aatccaatat	taagagatga	ctaatataga	45540
aactttgaaa	gtgagggaca	aggaacaact	aataaaaaact	acttagatcc	cagttcgcac	45600
ctgcagaact	taattttaaat	attttccaat	cgcaaaagtg	atatatcaca	cagtcctaga	45660
atatacttct	gaaactagat	agaatcttcc	tgctggaaat	tgaaactggg	tagaatcttc	45720
ctgctagaaa	ttctgattcc	tcttagtaac	ggagacaggg	ctgaggaaca	agggcaccag	45780
ttacgaagac	gacaatttaa	aaaggagctg	ggaagtcacc	tgggctatac	ttcaaaataa	45840
gtgcatgctg	gagaggacta	caaaggggtg	agcagtaggc	aaagaggatc	aggctaactt	45900

78

79

tgcccaaata	cagttatgta	tcacttaagg	acagagacac	attataagaa	atgtatcttt	52800
agtcaatttc	accattgtgg	aaacatcaca	gaatgtattt	acacaaacct	acatgggtata	52860
gtctactaca	tactgtggct	gtatcatata	gcccgaatgct	tctaggatac	aaacctgtaa	52920
agaatattat	taaactgaat	gtcgcaggta	attgtaacac	aataagcatt	tgcatatcta	52980
aacataccta	aacatagaaa	aatacagtaa	aaatatggta	taaaagataa	aaatgggtata	53040
cctgcataag	aggcttatta	ctgaaggctt	gcagacaggg	aagttgctct	gggtgagtca	53100
gtgagtgagt	ggtgagtga	tgtgaaggcc	tggaacattac	tttgactact	tttagacctt	53160
ataaacactg	tgactttaga	atacactaaa	ttaatTTTTT	aatttctttc	ttcaataaca	53220
aattattatc	agcttactga	attttttatt	ttataaattt	attataagtt	tgacattttc	53280
ataacagctt	aaaaacattt	tacaactgta	caaagtattt	tctttgtgtc	attattttgt	53340
aagcttttgc	ttttttcaat	ttttttttta	gctttttcaa	cattttgtta	aaaactaaaa	53400
atcctacaca	ttagtctaca	tggtcagaat	catcatcatt	actgtcctcc	acctccacat	53460
cttgtcccac	tagaaggctc	tcaggggcaa	taacatgcat	aaagctgtca	tctgatctga	53520
taacaatacg	tcgtctggaa	tacctaactga	aggacttgcc	tgaggatgtt	ttacagttaa	53580
catatataat	ggtacatata	catattgtat	acaaatacat	aaatacatca	tatacattac	53640
attataatat	atttatatat	tatatatata	acatatgtgt	gtgtattttc	atttatgtat	53700
ttattatgct	atactttctg	tcattttaca	gttaacatat	atatgtgtta	tgtatatata	53760
taatacagtt	aacatatata	tttgttgtgt	ctgtgtgtgt	ataaagtaga	aagaggataa	53820
tctgaaataa	tgatagaaat	tattgtatgg	taaatatgta	aaccagtatt	aagttagtta	53880
tttattacca	ttaccattac	caaatactat	gcactgtgca	taattgtgtg	tattataactt	53940
ttaaacaact	ggcagcataa	taggtttgtt	tatatcgaca	tcaccagaaa	gatgtaagta	54000
atctgtttgtg	ctaggacggt	agaatggcca	caacatcact	aggatgatggg	aatttttcag	54060
ctgcattgta	attttatggg	accaacgttg	tacatcattg	actgaaaatg	tcattaggca	54120
gtgcatgact	gtattttggt	aaattccagt	tcagatgttg	ctgtgaagg	attatttaga	54180
tgtgattcac	attgaaatca	gtagtttttg	agtaaacagt	ttacctcta	taatgtagtt	54240
gggcctcaac	cagtcagatg	agcactttaa	agattgactg	aggttcctct	aggaagaaga	54300
gactcaggct	gcagactgcc	tttggcttca	agactgaatc	ataaaactctt	ccctggctcc	54360
ctaactacca	gtcttccctg	cacattttgc	actttccagc	ccctacaata	tcacatgatt	54420
caactctctt	ttccctttct	ctctttcatc	tatctatccc	tctacctatc	agtcagcatc	54480
tatatgccca	cctattttatc	ccgttggctg	tgtttctgtg	gagtaccttt	actaatataa	54540
gtcacaaga	atagcttgct	ccaaaccag	aattcaggta	ctcaggtaact	atatatatcc	54600
caaggctctg	aagctgactt	tgtcctattt	tattgtaact	tcttaagaac	tgggagttca	54660
ttcataggg	ttataaaacc	ctacagattg	cattatttca	tggctctattt	tatatagttt	54720
aactgttctg	tgaatataat	gtttgatatt	ttactctttt	gatttttata	cagaatatta	54780
tataatata	attctggaaa	tatagttata	tatttccagt	aggaaatgtt	gattgagcat	54840
gatgcagaaa	attaacacat	gaagtgtctg	gattttcagt	tcattattct	cagtgcagtg	54900
ttcagagctg	tttctaggct	gcagggtcaa	tcatttgaaa	ttggtaattt	tagataatag	54960
ttgctaaaaa	gctcttccag	atgtaggata	ccctgtgtat	ttgaggaaatg	tacattccct	55020
gacatacaga	agttgtctac	tttttcatta	tatatggctt	gcattcttta	ttttccatcc	55080
tgccctgccc	tttactgcct	ctttgtctc	tcccttccct	ccaccattta	ctaattgggtg	55140
cataaatcaa	aataatttcca	aatcttacag	gtttcagggt	aaatgcaaca	tctgtcgtga	55200
aatctttctt	aagttccctt	aactaaaaat	aatctctact	atttctattt	ctctttatac	55260
cccttctaca	ttttattaat	ttccaatttt	atgatgtata	tgttcatacc	aagaaggttc	55320
cattttctgc	tttgtaaacc	aagtacatta	taaaactaac	aagaacactg	aaaaatgtct	55380
gctaaaaatt	taactgatga	atgcacacga	tgtatgtgtc	ctttggcaat	ctacattcat	55440
agatataaaa	agagagcaca	aagggttagt	agctgaatcc	agaatatgtg	tctacttttt	55500
gaatctctct	cctgtattgc	tgatgacttc	aatgatttaa	gcaattggca	taacctcttt	55560
gttccctatt	tctttttcta	tttcaaacad	agaataacac	caagttttgt	atttacctct	55620
tctgtctaga	attcagaagc	tgggtgctttt	acaataatta	agcaaactgg	agctgatacc	55680
tataaaattt	taccactaaa	ggtatttaat	tgtcctatgg	attagctgta	cttgactttt	55740
ttctctacat	gtctatagct	ccattactgc	taaaaattat	taggtccacc	gttctcttct	55800
ggccatgtga	ctagacagtt	ttggaagaga	ttaatgccaa	cataagaata	ataaatcagg	55860
atcaatcaaa	ttttatgatg	aaattatgca	cctagggttaa	ttgctttggc	aggtcttgaa	55920
ttttttgtga	agtttctccc	attactagaa	agacagtaat	attgaattaa	ctggtaattt	55980
gaaatgtgag	atacagtgtc	taattgcaaa	ctgacaatgt	agtaatgaga	ctgcactaca	56040
cacttatcaa	ttattatttg	attttattaa	gttaggaatc	aaaatggctc	acttcacatt	56100
tattcactaa	tgctgttttg	tcagctaact	tgcaaggaga	tacagcacat	tttgtaaagt	56160

```

acattggatg tcaaatttta aatgcactgt tccttatctc aaaatgttta tcttagaaac 56220
tcttgatatt togttctgag taacacagaag ttccatttta tgtttttctt ttgcttattt 56280
cttggtgtaga tcattataac ctgtgaacag gaacatgaag tacaagagca gtttggtcgt 56340
ttgctggaaa ttacactaat acaaaactgg attttttttt attacgtgta acatggcata 56400
gattttaatg gtcttcttat ttcagtgaag catttattaa aaattagcaa tttaaaagg 56460
cttataagta ttagcttctg caaagggtt gggatggatg gattatcaag acatatgtc 56520
tgagtccttc tcttcttgat attagattga ggggacaact gtgagtagaa aactttttcc 56580
atttctgatc ctttggaagt cataagtcct ccaacagtga ttaacacttc cagaatcagg 56640
gattttaaag aaggctctgc agctgctaag cagtattgac agctgctttg gaaatcccag 56700
ccttagacaa attggttctc aagtgccaca tttaatgcc tgtgaattag gaatactgca 56760
gctaagtata actgccagt agcattacat tccctctacc aaccagtgtt gaatatttgc 56820
ctagccgcaa tccttactgg tatcttaagt agtttctgct gtttgatcct atgcaacagt 56880
acaacataat tctggcaaat aaagttgact aaatgctctt tccagggcat ttagtttacc 56940
atagtagagt ttgagccttc tgtctatttg tatttgccat taaaattaga catcatgtca 57000
ttagcgtagg ttacatggtt agactgaatg gttagtcatt tgtgtcaggt gaagaggctg 57060
cttggtctaaa ttataattga ttccaaagac aaaacatctg gcctgagaca caggcaggca 57120
gagtggggat taatgataac ttcagtgaag ttgtccaaga atttttcaca aggttcaaag 57180
aataggtaca ccagtgatgt tgtgtataag ggtctcttgg acctcctaaa tgcattgcaca 57240
atttatgtgt tactccatta atattactca ataagatgat taaacatcta cagcgtaaat 57300
acgaattact atgtattatg ggatgcggtt tgccttatga ctacagcctgt gctttaaaca 57360
cagtataata gcaagggtgt ggctatcaga tttataaggc ttacatggtt gcaaaagtac 57420
tttcagtcca tattccacag cacagtttga tctgtgccat cccatattaa gttgcaaaaa 57480
ataaaatttt aattatctcc ctaagttact ctagcacagc aatcaaaatt atgtaaagag 57540
agggtcatga tgcaagcatt tttgtgcaca aaaacttcat attttaaato cattcaaaac 57600
aatgactgac ataatataga tgttaaatat ataggtagac tattgaatga ttacacatta 57660
cttcattata tgaataaatt attcagaatt ataagtgat atatttttaag ccctttataa 57720
aacatacatc tcatactcaa ttacatagtt tctggtacct attggatttt aaataattat 57780
tatgtatatg cttttccaga attgtaacac tgttcatcaa aaatagataa atgaggttag 57840
cttctcaatt tatttttata ttccagttta gcaaattgta attataagta gtatgcctgg 57900
aacctaggat ttgatccaac ttccctataa ataagttgaa tctagtttgt ttacttatga 57960
accacttgta attttttttt ccatttttca tttattggcc acttccocaa ttgtacttcc 58020
atttcttata ttaaataatt tccttcagga aaattggcca attgctaata cgacatcaag 58080
tgcaagtaat caatctctcc caccagaagt aacttatata atacctattt ttcaattttac 58140
ttattctttt aaatatcaaa agtatataca tacacatgca aacatacata cacacacata 58200
tatacacaca catatataca cgcacacata aatatatata gtacatatat ataatatata 58260
tgtatattta tttcctttcc tctatggcat tgattactgc tgcattcttg ttcattttat 58320
atatgttaaa tgcctgtaag atggcagaca ttttgtcagg atacaaagggt agataatatc 58380
taatcatcta ctcttgaaag ttctctgtta attctggaat ttctaatta cagttactat 58440
ggacaaaggc ttgtctctaa ccttggaaac catcagactt cactataacc accattatat 58500
gcatttggtt tggagtgtga cttgggattt agcctgcagt ggctcagttt ccatatcaac 58560
tgcttattag ttctgtgatg tgagaagaat tattttatat ttctattaat agatgttact 58620
gaaaaataat gattatagta gtacatatat cacagagttg caataaagat ttagtgagat 58680
gaacatgcaa agctcttaac atatgttttg cattttataa ttgctcagta agttagtttg 58740
tggttgcttt aaaacattta aaaacagtag atataaataa ttggaattaa ctttttagac 58800
aaaaaacaat gtaagggtc ttaaaaaatt tggttgtttt ttgctctag tttctttttt 58860
gtttatgtga cttaaactgt atgtataaat taatataatt gttgacatta agataacttg 58920
cttttaagtt taatgagtac atoctacata ctacaaaatc ttgagtaaaa ttcacacaca 58980
aaaaataaaa aataataaag ccaactgtgc aacaataact acttgaaatc ttggattttca 59040
ttttacatga ctcatgcaaa cagaaagaaa gtttaggccg tgtgcggtgg ctacgcctg 59100
taatcccagc actttgggag gatgagccgg gcagatcacg aggtcaggag atcgagacca 59160
tcttggtctaa cacggtgaaa cccctctct actaaaaata caaaaaatta gccaggcgtg 59220
gtggcggtgt cctgttgtcc cagctcctcc ggaggtcag gcaggagaat ggctgaacc 59280
cgggaggcgg agcttgcggt gagccgagat tgcatactg cactccggc tgggagacag 59340
agcgagactc tgtctcaaaa aaaaagttaa aaataatggc atttttttcc tccgaaaaga 59400
ctgagtttaa ctgttacttg cagaaacttt acaatacaaa aattcaaaact ttcgagaaaa 59460
acaactaaaa tagtagcttt actaatatgt tgttactctt cttatatatt aggaatatac 59520
aatagatata atttttgccc aaacctgggt tgacagtcgt ttaaaattca atagtacat 59580

```

```

gaaagtgcctt atgccttaaca gtaatatggt tggaaaaaatt tggattcctg acacttttctt 59640
cagaaactca agaaaatctg atgctcactg gataacaact cctaatacgtc tgcttcgaat 59700
ttggaatgat ggacgagttc tgtatactct aaggtaatat tgggtttaaaa atctgtagtt 59760
gctctcattt attttccata ctaatattta atcaattctt aattttttat ttctcaaagt 59820
tgcaaaacca aaaatgtgga gctcctttca catgctaaaa tataaaatgg tgaaatactt 59880
tggcaatact tttttaagct cagttattat tatgtgaaaa ggactaaaag caaagagaaa 59940
aaattctttg attttcttcc tttgaagcag aacaggttgc agcataaaga ttatggtttt 60000
cttaacctgc agtgcaaaata attgagtaag atttcctcta gaaaccacc ctgctataaa 60060
attttataat actctaattt tcaatggtgt ggtaaaacat ccctaagaaa ggattaaatg 60120
aaggttgatg tataaggttc aagttttatt aactcattga cttaaacata tctgcttatt 60180
aagaaccatc cttcatggaa aagatattca taataaaacca agagaaaaaa ttaccatact 60240
taaccattta taatctgtta atctgctatg tttgcatatg tgtgtttgct gacatattga 60300
agtatgaatt attatgtgta aacacagaga tcatattgaa atgtatacag aggtaacct 60360
gttgactgaa gaggaaattg ttagaattct gtcatatata aatattataa tattattgct 60420
atttaagact ttgagacttt agaatcatta ttatattcta ataggtcaca atctattatc 60480
atatgacaaa atcttgtttt aaaatgagac aggagatttt taatgcttca aattatttat 60540
ctttatttaa tacttttctt tccatctaaa tagattgaca attaatgcag aatgttatct 60600
tcagcttcat aactttccca tggatgaaca ttctgttcca ctggaatttt caagctgtaa 60660
gtaacaaaac acttgataat ttcatagaat ttaacctta gaagctgttt atttaaaaat 60720
gaaaagaaaag aaaataaaat aggattttca aaagtgtgaa ttatgttcaa ttcttttaga 60780
taaaataatt acaaataata tatttgcttc ttgcttttaa actaaaagtt tacaataaat 60840
gtgttatgaa tgcataatatt ttattaaaac attctttata acttttaatt ttattccaat 60900
gctatttgct tatttttatt aaataaagga gtattgaaaa taggtttgta ttcaaagtgt 60960
tttacacagt ccccaagaaa aggtgattca aattatattc atcttcctgg aggcaagatt 61020
taactttatt acattagatt ttgaaatgtg ttccaataat tgggttaggca aatagtattt 61080
accatacgaa tagcaaatat ttgagactac tttaataagg tagaaaaacc aaggtagaaa 61140
taagttgatt taattaaaaa agattgattc atagagaagt ggtagagggt gaataattag 61200
agtataaaaa ccgggtatat tcttatcctt tgccacttt ttggtggggg tgtttgtttt 61260
ttcttgtaaa tttgtttgag ttcatgttag attctggata ttagcccttt gtcagataag 61320
taggttgcaa aaattttctc ccattttgta ggttgctgt tccactctgat ggtagtttcc 61380
tttgctgtgc agaagctctt tagtttaatt agatcccatt tgtcaatttt ggcttttggt 61440
gccattgctt ttggtgtttt agacatgaag tccttgcca tgccatatat ctgaatggta 61500
atgccctagg ttttcttcta gggtttttat ggttttaggg ttaagtcttt aatccatctt 61560
gaattaattt ttgtataagg gataaggaag ggatccagtt tcacctttct acatatggct 61620
agccagtttt ccagcacca tttattaaat agggatttat taaatagggg aaaggatttt 61680
cccattgct tgtttttctc agctttgtca aagatcagat agttgtagat atgtggcgtt 61740
atcttgagg gctctgttct gttccgttga tctatatctc tgttttggtta ccagtacct 61800
gctgttttgg ttactgttag cttgtagat agtttgaagt caggtagcgt gatgcctcca 61860
gctttgttct tttggcttag gattgacttg gcaatgcggg cttttttttg gttccatatg 61920
aactttaaag tagttttttc caattctgtg aagaaagtca ttggtagctt gatggggatg 61980
gcattgaatc tataaattac cttgggcagg atggccattt tcacgatatt gattcttctc 62040
acccatgagc atggaatgtt cttccatttg tttgtatcct cttttatttc attgagcagt 62100
ggtttgtagt tctccttgaa gaggtccttc acgtcccttg taagttggat tcctaagtat 62160
tttgttctct ttgaagcaat tgtgaatggg agttcactca tgatttggct ctctgtttgt 62220
ctgttattgg tgtataagaa tgcttgtgat tttgtacat tgattttgta tcttgagact 62280
ttgctgaatt tgcttatcag ctttaaggaga ttttgggctg agacaatggg gttttctaga 62340
tatacaatga tgtcgtctgc aaacagggac aatttgactt cctcttttcc taattgaata 62400
cccttcattt ccacttctca aaagaagaca tttatgcagc caaaagacac atgaaaaaat 62460
gctcaccatc actggccatt ggataaatgc aaatcaaaag cagaatgtga taccatctca 62520
caccagttaa aatggcgatc attaaaaagt caggaaacaa caggtgctgg agaggatgtg 62580
gagaaatagg aacactttta aactgttggt gggacagtaa actagttcaa ccattgtgga 62640
aatcagtgtg gcgattctc agggatctag aactagaaat accatttgac ccagccttcc 62700
cattactggg tatataccca aaggactata gatcatgctg ctataaagac acatgcacac 62760
gtatgtttat tgcggcacta ttcacaatag caaagacttg gaaccaaccc aaatgtccaa 62820
caatgataga ctggattaag aaaatgtggc acatatacac catggaatac tatgcagcca 62880
taaaaatgtt gagttcatgt cttttgtagg gacatggatg aaattggaaa tcatcattct 62940
cagtaaaacta tcacaaggac aaaaaaccaa acaccgcatg ttctcactca tagatgggaa 63000

```

83

```

gcattgcaat ggtatcaaat ttggatcttg ctaatctatg ttttttttct ctatagatgg 66480
ataccctaaa aatgaaattg agtataagtg gaaaaagccc tccgtagaag tggctgatcc 66540
taaatactgg agattatatc agtttgcat tgtagggtta cggaactcaa ctgaaatcac 66600
tcacacgata tctggtaaaa agaatactca aatagtgaag tgatgctagc attaaatgac 66660
atttttatct taaatatatg gattgtgatt tatattatat agccaaaata aaactttttt 66720
ttcttacagg ggattatggt atcatgacaa ttttttttga cctgagcaga agaatgggat 66780
atttcactat tcagacctac attccatgca ttctgacagt tgttctttac ttgggtgtct 66840
tttggatcaa taaagatgca gtgcctgcaa gaacatcggt gggatgaca tgtaatatta 66900
tgctataatg ccatagaact ttaaaaaatc attttgatgt gataaaagt catagtcaca 66960
tttattagat atattcttct agactacagt gcagggaaaa aaattgacag gaactaagg 67020
acaataccat gttgaaagag atatgtaaaa aaagaaaagt agaaaatgga attgtaaaga 67080
acttgaaatt tcagaaaatt ccttcttttt tctgctaatt aagaccaagg gaagcataga 67140
tccataacaa gaattaatac atgaaaatca tacgtaattc ctaaatctta ttttgggtga 67200
gtttttctat cacgtttcat caagatttta ttttatgtta tagctgagtc actaataacc 67260
actgtttatt tactgagggtg tgcagatggc tttaaaataa cctgctttta taacattgct 67320
aatgaagtca atctttccaa agggcaagta gctgaaagga acaagacgct ccctggcttc 67380
tgtggcttct gtcagtata aaatgtgtat ttattctgtg aatacttgag agttgcctgt 67440
agttgcctgt agaataacat tgattaacag ggtaatttga ttcatacttg acaaaataga 67500
cttttactat ttttctgtaa caatctattg taataatcaa atcacataac aaaataagtg 67560
actaattaca gtatagttat ttaaaagaaa atgataacct gcaatcaggt attgatatag 67620
gaataaaaact tgttaatcaa ttgattcacc ccccaaatca gtgaaagaca catgaatcac 67680
actgctcata attgtgtctt aaaaatatgt aaagaaagggt gagacagttt ctatttttaa 67740
ccacaaagaa gtttatatgt gatttctatg atagcttata aatgattcca tctaataagt 67800
acagtaaaat gcaggtcaca taaaagtgc aacaacaatg catgaggcac atgggtctat 67860
acttgaaatt ctcttgccca tctaaacagg gatattgaca tagctagtgg acaaaatcca 67920
ttcgtatctc aattttgttt cctttcattg tttaaaatta ttttataagt tgtataaatt 67980
atttgagggt ctttgatatg tactataact taaccatagt atttttgttt ctatttccat 68040
atttttctgg gaatgtaaat tagtataacc actatgcata actgtttgga agtttctaaa 68100
agaatttgaa tggattatct ataccactgg gtttgataaa agttctatgc tagagtgtca 68160
tgaagagaaa agccaaactgt gaagtcgttt ttgcataaac ttatttttat tcaatattag 68220
catatcaatc ataaaggtaa gcacagccag atgtaacttg acaaatatac tcttgtaatt 68280
gtggacccta agatagagaa ttacaattgc tttaaatata catgcacaca catatcactg 68340
caagacaatt tctgaactc aaccttgcat ttttctacaa tgtatccaat gctttctgtt 68400
aaacaagcag tgtttttggt gcttttagtg taacagtga taaaatttta aaacttcgtg 68460
tttatgggtg ttacatttta ttagtagtaa aacaggtaat gtacaaatgt aaaaggcata 68520
caataaaaatt ttagtgagtc atgaaggctg tgaagagata gcaaaatcaa tgtttagaaa 68580
atgacagggt gtgagtggtg tgatatttgc tattttatat tgtggagtct caatagatac 68640
ctctgatgat gtggcatttg agcagtgacc caatgaagtg tgggtgtaac ctggtggata 68700
ttgggcagaa agatattcca ggcaggggga accatatggt tttgagcc agatcttct 68760
gctatgccat gctttggggg tcaagtaaaaa gacaaatgtg gttggagctg aatgaatagt 68820
cacagcaaaa taaatttgga gaaggggcca atgattagat caggtaaatt tttgtgggcc 68880
attttaagga ctttaaaatt tatttcaaat tagtgaggat actatttag gatcttgagt 68940
aggggaagcc atggagaata tagcaatgat atccaaaatg ttttctcaat atcctgttaa 69000
tatcttaagt gttttttatt atcacggact cagaaatatt ttaatagaaa atccactatg 69060
ttcttctcaa ggtaatgaca gtgataataa ttactaaatt aaataaatgt ctgaaagcaa 69120
agttctgctt agtttttttg ctttattaaa aaagaaagta acagaaactt aataaagtaa 69180
aatttattgc agtagtggt aacaaaatga aaccacacat tggatttaag atttttcaga 69240
taggagacat aattgtggag ccaagaaaca ctttggaac ctacctgggc tgaagatggg 69300
aacaatgtga ttttaataagc ttgtgtggga gatgaggcaa agggagactt gtatataatca 69360
tttttttagc ctttaccata acaaaagcct taaaaatgta catttagttt caaaataatt 69420
gaagtaaaag ctgtaaggaa tagtgggaga aataaacaaa tctgcaacca tactgttttc 69480
tcattcattg ttagaactac aagaaaaagt taattaggat atagaaactt tgaaaagaaa 69540
ttatcatcaa ctactttacc tgattgatat ttataggaca ctacacctaa cttcagagca 69600
cacaatattt tcaagtgc atagatcatt aaacaagata ttactaaaat acaaaacact 69660
aaaattttca gagtatattt tctaacaaaa aaggaagtaa attaaaaata aataaaaaata 69720
atataactga caatggcctt gtatatattg aaatttaaaa aaaagacaaa taattaaatt 69780
aagtacattt tgtgtatatt tactacaatt acgaagcaag atattactgt gtttgttttt 69840

```

tttaaacgag	aggcacaaaag	gagttgaaaa	actttgttaa	gatcacagag	ctaatagaag	69900
agtcagagat	taaacctaaa	tttgtatcat	aattccaaat	tcatcacaac	tgtttcacct	69960
ttaatatatt	ctccaatcct	aagaaatgat	aaaagtgaag	tgacagggtt	gtgagtagaa	70020
tttttaggg	atgtttcaat	ggagacatga	ggggaagagg	tggttattgc	cttcactata	70080
ccaaagagac	tggttgctaca	gctgtttttg	atttactttt	catcataaat	gttgtcactc	70140
tataaatatt	ggtcaaatta	agaagcattt	agcattatgt	ttagtacttt	taagagagta	70200
aacaattccg	ctcatggggc	catattttact	gggaaaacga	atgaggatta	tgaaagttaa	70260
ggttaaaaga	aatattagca	tgtgcatctg	taatttagga	gacctaaagt	taatagtgat	70320
aagaaccaca	caaagactga	tgagaaacag	atcccttcaa	ggatacacat	acaggattgg	70380
gacccttggg	gacatcccag	gtatacatgc	tgacttctca	aaaataaaag	gattttactct	70440
gtgctttcaa	agcagtattc	ccttgatata	actgtgccag	tattcctcac	ttaaacattg	70500
agattaaaaa	atgggtttta	ctcaggttat	tatataacaa	aaagctatgt	tagtttggtg	70560
taattttgac	tgtatatgta	tgggtgaaaa	tgtcaagtca	ctgtcacttc	aattagatag	70620
aagaatatct	gttgcttatg	tgaagtctag	gggtagggtat	ttcttttttt	tttaatttca	70680
aggcgataca	atgcattata	gtatgagaga	ctcaggctcc	ttctggcttt	ctcctgtatt	70740
gtgagtaa	tttaattcca	atattatcta	atgttttaag	atagtgtctat	tgggtctggcc	70800
attgtattcc	agcagaaggc	aatgagaagg	aaagaagggc	attccttgct	tctttaatgc	70860
tgtctttagg	atattgtgtg	ttgcctacat	tcttgacttg	gttccctttg	gctagatctc	70920
agcctgtagc	cacactcaac	tgcagagaga	tggcaaataa	agttgttctg	ggcagtcata	70980
tgccagctaa	atatttgata	ttctattaca	atggagggaag	ggaagaatgg	atatcaggag	71040
tcaacttgca	ctccgcagta	atttaaatga	ttagaactca	cagaaacaca	aagagcacat	71100
tggaacacat	gttgctatgg	ctaccaataa	ttattaatat	aaattactag	taatcatatt	71160
tttgtttcta	ttgtttaatt	tcactcacia	tattggccaa	tcagatggaa	gaataaatgt	71220
ggctatttaa	ttttcatttc	ctgttttggt	tcctctctct	ataggatatca	ctacagttct	71280
gactatgaca	accctgagta	caattgccag	gaagtcttta	cctaagggtt	cttatgtgac	71340
tgcatgggat	ctctttggtt	ctgtttggtt	catttttggt	tttgagcct	tgatggaata	71400
tggaaccttg	cattatttta	ccagcaacca	aaaaggaaaag	actgctacta	aagacagaaa	71460
gctaaaaaat	aaagcctcgg	tatgttaaaa	agaaaaattt	ctattttata	aacctcaact	71520
tacttagaga	tactcttttg	aatattttatg	tgtttgttta	cagaaatagg	gacgcttgag	71580
accataaacc	aaagacaaag	ttaagatgaa	actggagtac	ctaacattgc	tcaccaagaa	71640
aagtgtat	ctgtttttca	aatgtctccc	attggatccc	ctgtgtaaag	agtgggttat	71700
tatggcctga	attttcccac	agtggtagat	tctttgtacc	tcaaacatta	tttgtgattt	71760
acctttgatc	cgattttttg	tgttaggatt	tgcatttcta	acaaaattac	cttactttta	71820
gtgatgaaga	ttcctcttta	atggtagtgt	ccaactgtgt	ttaaaagtga	tgaccacaac	71880
gttctttttt	tgtttttctg	tatttgtact	attttactct	cccatctcaa	tcataataat	71940
accttaactt	atatacacatc	aagtctaata	tattaaaatg	gattctcctg	aatctaatat	72000
actcatgtat	gggtttattg	actatgatta	tgttttatat	tatcaattat	agcaacctga	72060
aagtaatatc	taacacatgt	tgaatgttcc	agacactctg	cagagcattt	tacatgcatg	72120
atttcattga	tgcttagagc	aactctctaa	gtagaggttc	aatgatttac	acaagaacag	72180
cctcagtaag	cgaaagatcc	aagactcaga	cctggatcct	ttttgggtgc	agagctcaag	72240
cgtgtgaccc	ttcctttata	tatcctgttc	tcacagggtc	taaaaaata	gacattttat	72300
tttctaaaa	taacaagaaa	acattttaatt	tagtaaatgc	tttcaaagat	ttgtttcaaa	72360
atactcaaat	aatgagaaaa	tgtatcaaat	agccaggcat	tagtggcaag	cttghtaagcc	72420
taccttaatg	tatatataat	gagttttcca	tattatctaa	ctttttaagt	tctaggagtc	72480
aatgcatttt	tccctaattct	tttttcaattg	gtttctttat	aaaactctgc	tatactttta	72540
agttatgttt	tacttttcat	gttatattcc	aagagaaaaa	cctgggtgat	tttgctttca	72600
gatacccaca	atattccaag	aaagcagagc	acattgctag	cagtacaaat	ttgataactt	72660
gcattgtaga	attagccagt	gactccgtca	acttcacgtc	atttttgctt	atctatagga	72720
tatatattgt	ctcctgaatg	gacaaatcag	aagtcaacta	taatatccat	tgttacattt	72780
ttagataaag	aaaaatgaat	gctttctaca	attttccctt	catcaaagcc	atgcatcttc	72840
tctggatttt	ccatttttat	gtttacttat	ctgtatctat	gttggattat	tgcttgaaa	72900
gttttttttc	tggaaatttt	gagtgagat	gtttagaata	aaaataacta	atatagaggc	72960
caggcaatgt	accaatctac	atccattctc	tattttcatt	ctaagaataa	ataatttagg	73020
taccactatt	atattatcga	aggacacaac	caaaacatgg	ctactctagg	acggtagcta	73080
atcagtttga	ccgcagatca	tgtacaactca	gccaaagttc	caatggcaca	cttactcaag	73140
agagctggat	tactctagtg	tagactttta	acttatcatg	ttattcaatc	attcatcaac	73200
tattttattaa	ggacttccat	aatctaggca	ttgttctaga	atagttacta	tttttaatcc	73260

T0301067.03301

atatgttaag	tctttcaa	ttagcatcaa	gcctaggttg	caaactacta	attatggaaa	73320
ggtaaagaac	gaaaattaaa	tgatagatatt	gaagtactaa	atggcaaata	gcaacagtta	73380
aattgacata	ttataattga	ctacttgagt	acaaaaata	ttttctttac	tgccattcat	73440
aatgaagttt	gatagtctct	tttctttggt	caattatgtg	ctttatgagc	tctttcagtc	73500
ctaaatgata	ttttgagtga	aaaaccggga	tcagtaataca	aatctgactc	cttgcattgt	73560
tcagtgtaca	gccatatggg	cttgctaatac	aggtaaataca	gaatgtactt	ttatgaggaa	73620
taggcaactc	tgtttgttgt	tgtctcaaat	cctctatttta	agggttttct	ggaatccaca	73680
ttttcttttg	gttgcaagat	gttgtagcat	cttgccatta	agcctgtatt	aaatgtttgt	73740
gacttcaccc	cctccttccg	cctttctctc	tcctttcctc	tttgtcttcc	cttcaatctc	73800
cccatttctt	tattaaaatt	cagagttagtt	taaaaatccag	cattttccct	ttaaaaaaa	73860
tgtttagggc	tatagtcagg	ctaggatgca	atactttggc	tagagagttg	ctgaaaggtc	73920
aacaattctg	ttattaagca	aaattaaaat	ccttttcctg	acatcttggt	gactgcttaa	73980
ctgagaatct	caaaatgttc	agtttttttt	ggaagagata	atacttgaac	aaaatagatt	74040
tttagcttca	aagaaccacc	taaaaatata	aattgataaa	cacccgtttt	aacttactct	74100
ccacatgaaa	tgggccagct	tgtttgtctc	ctctcactct	ttctttcatt	ctctttcttt	74160
catttgcttt	ttatgctcgc	tttctatttt	tccttccttg	cttttccttt	ctcccttccc	74220
tttcccttct	tccttccttc	tttcttttct	tcctttcctc	cctttccctc	ccttcgcttc	74280
ccttcccttc	tccttccttc	cttcccttct	tccttccagt	aagaagggtta	cttaaacat	74340
tgtaaagagt	cgggtgatgg	tgggatattt	tcaatagata	gagtaaccat	ttagtaccag	74400
ttgtgacaat	ggccatatga	tgttcccttc	gaagtatata	gataccacca	ccaacagctc	74460
ttttttctac	cttggagaa	ccttcccttc	ttccttcctt	ccttcccttc	ttccttcctt	74520
ccttcccttc	ttccttcctt	ccttcccttc	ctattccatt	atgaccaa	ttttcttcca	74580
ttttattgta	catgtgaatg	acatttttga	ttgtttcata	tgcccatcca	atatatatta	74640
ttactaatcg	gcaagtatat	aattcaggtg	tatttgtgcc	aacatttatt	ttcaagaaag	74700
tctaagaagt	taagttaaat	atgttgaact	atggaatcaa	gacattttat	gcaacagata	74760
aagccactag	taataaacog	aaaggcacta	gtaataaaact	agcctttgcc	aattgaagag	74820
tagtttatgt	tgaacttga	atattcggcc	cataggctag	cagaaatttc	agaaaaccat	74880
gttgtagcta	cttcaggggc	ctgctatagc	catttgagga	ttgaacttac	ttatttttag	74940
tttctgtatt	ttattaagat	gtgaaattat	agaaatttat	ttatccagct	atggacgaga	75000
cattatggag	acaaaaaaga	gatagaatgt	gttctttgct	ttcagggatg	attacatcca	75060
gttgtagaaa	aatgcacgc	aaatacaaac	ataataagat	acaatataaa	taatattcca	75120
agataaagat	caaaattaaa	caaggagtct	gtgagggaa	taacaattag	atgaataatt	75180
tctcatgctt	tttcagaagt	agggtatttt	tataacaaat	atacataaga	tattttatta	75240
tttttaata	acaatatatt	tattttttat	aaaaacataa	aatgtatttt	taaaggaact	75300
gttcttgtgg	aaacagacag	gtatcacaaa	gccagacctt	tgtggcctat	gcctagcctc	75360
ccaggcagtc	ttcagacatc	atattatgtg	gcctgtgaaa	cattgaacag	gcacacagtg	75420
ttgaagagaa	ccttgtggag	aggagaggat	tcctttcatt	tcattgtgta	ataattcatt	75480
cataaaacat	ataaactgat	ggaagtgttg	caaggaaaat	tcagagaaa	gtgtggcttc	75540
actgaagtaa	tttttgaag	aaaaagtgtc	agttttcaag	ataacaaatt	agaaagaaca	75600
tttaagatag	agggaactat	gtcagctggg	ctgagcctta	catgtgagaa	caaaaagtag	75660
actagcatat	gtcccaagct	gtgctgtcca	atataatagc	cactaagcac	atatagatag	75720
gttgggttaa	attaattaaa	attaaacaaa	atataatatt	cggatatctt	gtagcgctag	75780
tcacatttca	agtgtcfaat	agccacatgt	gacttatggc	tatcatattg	aagagtgtag	75840
atgtagaaca	ttgtcatcac	agaaagtgtc	attgggtaat	gtttatcttg	aaagtaggaa	75900
agaaacattt	tctcttagga	aagaaaataa	cctatgttga	ggatagaaga	gtggtttatg	75960
gccagggtgt	gcaaaatctt	taaagtcatg	ctaagaattt	acattttgtc	ttgttagcgc	76020
aggcaagtta	tacacgtttt	taagcaggat	ggtatatttc	tatgtatgtt	ttcagaagat	76080
aattagaaat	agtatccagt	atatacttgg	atggcaaaag	aggcaggggag	ataaatttgg	76140
agactattgc	agttgttcca	gaagtaatat	aatactaaaa	gctgaaacca	agatattagg	76200
atgaaaataa	tgcattagca	agtttttggg	tcctccctgg	caccgcctca	ccccacagtc	76260
ctcaccacac	gtccttttct	agcaaagaga	ttttcatttc	agtcagtgc	tgccatttga	76320
ataaacattt	gtctacctct	cagagtacta	aaataatgaa	aagctgtatc	aatccacagt	76380
ttcacaaact	atagattctt	tcttataata	gtatttggca	aagctccag	gaaggattat	76440
gtgatattgg	tgaatgtgta	aatacagcaa	gaaggttact	taaacaattg	taaagagtca	76500
gtgcatgggt	gaatgttttc	aatagataga	gtaaccagtt	agtaccagtt	gtgacaatgg	76560
ccatatgatg	cttccctgga	gtgtatagat	accaccacca	acagctcttt	tttctacctt	76620
ggctaccaaa	gaacactgct	taatatattt	atttctattt	aaaacatcat	agactatcac	76680


```

agcagaaagg gaccagatta atcattgagt aaaattccca tttaaaaaaa ttaatttttaa 76740
cagtaatagt ggagcatggg tgaaactaat aaacaaatag aacaattcat aaataaacia 76800
atagaacaat tcataaatgt acaaataaag aaaattctct acctagaaca ttttgtocaa 76860
aatctcagaa tcaactacat tgtaaccatt tctgatttta cattcttgta attattattt 76920
tcattgtatt tctgattcct actgtaccaa atttccagggt atctataact ctttgggtaca 76980
aaagaaaaaa tagctcttct gttgtatcct tcttattctc tttctcagta ttttaactta 77040
ctattaatag ttattgacct ggtgtattgg ctcaactcctg taattccagc actttggggag 77100
gccaagggtg gcagatcacg gggtcaggag attgagtcca tcctggccaa catgggtgaaa 77160
cctgtctcta caaaaatata aaaaaattag tcaggcactt ggtaattttt aaaactttta 77220
ataatatatt taaatctcta cttctgaatt cattattgtg attaattcta gtgccaattt 77280
tattatcatt ctttgcagat aaaacttttt tcttttctta agaaactttc catcatcttg 77340
tttttttaca gtgaaaaaaa aatcttaagg agtcgtacat tttctgtgtt ttgtcctctt 77400
tcatttctct attacttaat attcttatgt tttatttttg agaatgttat caattgttca 77460
ttccaattaa tttctatttt ctcttggaa actactggta gttagatttt tgacttctta 77520
ggttcagcat gtattccaga ttttcttttt atcatatttt aagccatttt tccatactat 77580
gagcagtttt ttcaagattt cattttaaca cttttagctt attaaaaatt tacaataact 77640
ttatgaattt cagagagggtc ttattttcta actgttcttt gttcattttt ccttaagtgg 77700
atgtattact tctcaagta tttctgagtg gactaactag atttattttt tcatatttcc 77760
taaaatatta gaatcagtaa ttccagctgt tcatatttcac ttttctcctt tgtattgttc 77820
acttacctca aaacttcttc atgattgttc cccaaatagg cccggtaagt ttctgatgta 77880
actgtcaagt ttgcaaccat cgtattttac tttcatacgt gtgtttgtgt ggcatgacat 77940
ctgattcctt ccaaatatac acatatacat tctaaatgat aggatgatgg cacaactata 78000
ggacattcca gttccaacat caagagccct aacctttctc ctaggatagt gaaagtctct 78060
tctcattttt tagaagtat agagttaga gatgagtga tgccaaccgt cgtcatcttt 78120
actctctaca tgcacagtag gtgaaccata gttcaaagga gtgcagctga tctggagggtg 78180
gacatttagc taacaccatg ttttccattc tattcattac tcccattctt ccttctttat 78240
actgagtga agagttagt ccacagtcac gtttccattt actgagcctt tctgagatgt 78300
gtcttctttt ggtctgcttt atccattggc ctatttccat ctgttttttg tacttcagaa 78360
attgatcaga aattctaaat gctgattggc agctcctcgt caccaagatt cactcctgg 78420
aatttgtttt ctgtttttta agacaaagtt tcaactccat acccaggctg gaatgcagta 78480
gtgcaatcac agctcactgc agcctcaacc tctgagatc aagccatcct cccacctcag 78540
cctcccaaat agctgggatt gcaaacacat gcaactacac ccagataaat ttcttttgtt 78600
tctctttttt gtggatacag ggttttgcca tgttgcccag gctgggtctca aactcctggg 78660
ctcaagaaat ctgcccacct tggcctccca aagtgtctggg attacagggtg tgagccaacg 78720
tgcccagcct tctcctggtg attttttaat gttgatggta taatctacaa tttttgaaca 78780
tcccaactta aacaggattt tgtgaaggag tggtaaacag tgcttagttt ggcattttct 78840
catgcaatat tttaaatggg aaattaatat tatatagtca gtatttcata ataataaat 78900
catcaaataa tatttaccat tcatattact aatatttaat aatttttagca gatggtataa 78960
taattatcag atgggtataag taataactat atgtaaatac tgatttattt atcaactccc 79020
atltatttga tattagtata ttttgggtct taaaatata tttattagta tataactaga 79080
cagagatata tctttaatag tgaaaatttg atgaatttct caaaatagtg agctcagcta 79140
tattcttagt aatttttgaa atttttttaa agtgtatttt aaaaacgatt ttattttaagg 79200
tagaatattc tattcaatta tgttgctact ttttaaatg ctttaactga tttgtttttt 79260
aatgtttgct tttttgggga aggcagggtc tattcttgag aaactaaaag atgggtagca 79320
tagctgcact tcagataaat aagacttgga atgtatacat ttttctgatt aatttctcag 79380
tggttttaaag gtcagaaatc tgtgtttcat caatatttta ttttattgac attgtgttag 79440
atltttggct atltttctca gcatltatat gtccatttct gtttactttc attgttcaat 79500
gatgtgtagt ttctttcaat tcatgaactt gagttttgag gggatggggg ggagggagca 79560
aaaatctaga caactgccaa tttaaaatgt atttcccata gaactctcca tgatactctc 79620
accagagagt tcatttttta aataaacccc ttgcagtttt tgttcatttt agtgcactaa 79680
aaaatttaat catttctctc atgaaaaact aatatcctct gaaaaacaca ccatatttcc 79740
atltatgata atgcccgtttc atatatggaa tgtgttttct accacataat cttaaaaattg 79800
cctcctaaca atltttttaa atlaattttc attctgatta gcataagagc atatgctatt 79860
caagtatcaa gtgacaagga aatagggtacc aattcgtctc tctctttttt ttacatgat 79920
tttagtgctc cattaaaaat taggaatgta atcaactatc cccagttcca tctcaaaag 79980
atctgtcatt aacattttaat aagatactgc ctctgctttt tgtgatatag ctaagcaagt 80040
ttaatcctga acttacttgt gtttggggca tttctcctta tctgcaagggt gcacatttgt 80100

```

```

ctctaataatg aatattgtga cctgcatgat gctttaatct tttaaaaata tcactaaatg 80160
tgaatatttg ctcccttttg catctactgg atttttaata tttggtgact tggaatctat 80220
gtttcttttt agaactaaaa tggacatgaa tatagttaaa gttgtaagaa attttaaaact 80280
tttgtataca tttgtaagaa tgtttttcat aaaatattta attttggtga cctattttcaa 80340
aatttagcat gaaaacattt aacaaacatt tgcattggaa aagatgtaag agctgggttaa 80400
aacaattttt ataaaatttg cgctatatatt aatgttttaga gagtgaattt gctgaatata 80460
ataagttgta agattcacat cttgagataa cctcaaaaat atggtgaatt ctgaaaaacg 80520
aagaattgct aaagtgtgaa tactcatctt ttacttgcca gcttattaaa ggagtcatta 80580
acttttttcc agtaacattt ttatgtggtc taggactttg ttgttggtgc tattattatt 80640
gtttgtatgt ttatcttagc atgaatttta taatcaattg tgatataggc ctcatcagga 80700
ataaacagta acaatgacaa caataatgat gttacctact ggcaggtgat tacttgctat 80760
gtttcaagta ttttgcttag tgtttttatat ccatcatccc atttaataat caagcaaacc 80820
tacaatgttt ttattgccat ttttgtagag aagaaaacat agaagccaaa gaataatact 80880
ttcacttttt aattatttaa gtattattca aaatcaagta ttataaaaaac ttaatttaac 80940
atgttccaaa taaagctggt aataatagta gctattatta ctgagcatga tggagcattg 81000
atcattttac ataagccatt gaaatttact tatcatagct atcacagaaa ttaggtgcaa 81060
gtataattcc agctaagaag acagatacat tggtagaatg agcagaattt gaacaccaat 81120
tgttcccaag tctgagcttt taattcctac acatcctctt tcccaaatat ggtgccttgt 81180
gcttacccat cacaacccc ataatctttt ttgtcaattc atatcttcac aaaccttttg 81240
cattatcaaa atggcttctt tgctatttat totcattttt attaaagtcg ataaataata 81300
tataatagca aaaaaagatg tgtatagcac tgttaatggt ttaatatata atagaatata 81360
cagttcattt tatttttagaa ttagatccat ggtaaagat attttggtga gtcaactttc 81420
taaatgtata aaaatatatt ttatttttca ggggtatttc atttctgctt aatagaatgt 81480
aacaaatggt ctattacaaa gcaaattata atataaaaaca tgttataatt gaaaatactt 81540
gattttttga aatcaagatt attttcatta cctgtcagtc toctagagtt tgcgttaaag 81600
gagcaaatgt atctttcctt atgctacttt ttgatgtca aaaattcatt tattattgtg 81660
ctcagatgac tcctggtctc catcctggat ccactctgat tccaatgaat aatatttctg 81720
tgccgcaaga agatgattat gggatcagtt gtttggaggg caaagattgt gccagcttct 81780
tctgttgctt tgaagactgc agaacaggat cttggaggga aggaaggata cacatacgca 81840
ttgccaaaat tgactcttat tctagaatat ttttcccaac cgcttttgcc ctgttcaact 81900
tggtttattg ggttggctat ctttacttat aaaatctact tcataagcaa aaatcaaaag 81960
aagtctgact aaattcagta gaatcttttg tacttcagta acttgaagtt taaattttaa 82020
atgcagagag accaatgggt aaaatgtgaa tagtattgta actattttta ggcccttcaga 82080
agtaaatata gtagcagctt tcaggctaatt ttacgtgaaa ctgattagtt gcaaaatcca 82140
gtaggttaaa atactcacat atttttactt aaattttctt taattttact atatgttatt 82200
ataattttga atttttaagt tctatgattc atgtttttaa gatggaataa ttttaatata 82260
tattttgttt aaatataatc tataattggt ttgtaatgta agactaatta ctaatattta 82320
tgtagcaact tttgtgccag aaaaaagact gtttaattgt tttttcttgc ttttcatttg 82380
attaccctgc ttgaaatata gttagttgat aacataaagc catagttttc ttggattttc 82440
ttccaagata ttgtattccc aagaagaat tgatttattt ttaaaactatc agttactgaa 82500
gacttatgaa aagggtcaatt tttacctgtc ttttaatcca gtccattttc tgacacaata 82560
ttaaacagaa cgccagttgc atttactttg gtgatttgca aacttggaat gaagccacca 82620
gtcatttttt aagaatgcaa gatgaaaaaa tcaggtaaaa tctaaccatt ttattctctg 82680
cttcatagca tttataatgt atgatgaggt taacactgaa atattaaaat ctgcaaatgc 82740
acattaatgc aaattaaatg cagagagaaa tgattatttt tttctttatg ttttatgaaa 82800
gtgttgagct ccctaaaata aaaatattga ggaacattgc ttgtatttac tgatgtggtt 82860
aaaaaattat tacctggcaa actgagttgc aggtcacatt gaggacaaaa tcttaaatat 82920
gtgctattag cattgttt 82938

```

<210> 4

<211> 465

<212> PRT

<213> Rattus norvegicus

<400> 4

```

Met Gly Ser Gly Lys Val Phe Leu Phe Ser Pro Ser Leu Leu Trp Ser
  1                      5                      10                      15

```

89

[illegible]